MASTER EN SCIENCES MATHÉMATIQUES

Méthodes de minimisation de la plus grande valeur propre d'une matrice symérique

Canon, Marie; Lambert, Delphine

Award date:
2000

FUNDP
Faculté des Sciences
Département de Mathématique

Rempart de la Vierge, 8
B-5000 Namur Belgique

Méthodes de minimisation de la plus grande valeur propre d'une matrice symétrique

Mémoire présenté pour l'obtention
du grade de
Licenciées en Sciences
Mathématiques
par

Promoteur : J.– J. Strodiot

Marie CANON et
Delphine LAMBERT

Année Académique 1999-2000

Nous tenons spécialement à remercier Monsieur J.-J. Strodiot pour son soutien et son aide qu'il nous a apportés lors de l'élaboration de notre mémoire.

Un tout grand merci aussi à nos parents et amis pour leur confiance et leur présence tout au long de ces quatre années.

Résumé

Le but de ce mémoire est de présenter une méthode d'optimisation non–différentiable pour minimiser la valeur propre maximale de matrices appartenant à un sous–espace affine de matrices symétriques réelles. Nous étudions d'abord les propriétés du premier ordre pour la fonction valeur propre maximale afin d'obtenir *l'algorithme des valeurs propres approchées* qui nous fournit le minimum recherché. Ensuite pour augmenter la précision de la méthode, nous présentons un algorithme du second ordre en utilisant la théorie du $\mathcal{U}$-Lagrangien. Enfin, en utilisant la dualité, nous exposons brièvement une méthode qui est numériquement plus efficace.

Abstract

The aim of this work is to propose a nonsmooth optimization method to minimize the maximal eigenvalue of real symmetric matrices. We first study first-order propreties for the maximal eigenvalue function in order to obtain the *approched eigenvalue algorithm* that provides us the required minimum. Then, to raise the method's precision, we present a second-order algorithm by using the $\mathcal{U}$-Lagrangian theory. Finally, by using the duality, we shortly expose a more sufficient numerical method.

Table des matières

Introduction

Les problèmes d'optimisation concernant les valeurs propres sont présents en mathématique depuis 1773. A cette époque, Lagrange a posé un problème qui demandait la maximisation de la plus petite valeur propre d'une matrice. Depuis, ce genre de problèmes traitant de l'optimisation des valeurs propres est devenu un domaine de recherche spécifique. Pour une description plus détaillée de tels problèmes, nous renvoyons le lecteur à l'article de Lewis et Overton [13]. L'objectif de ce mémoire est de résoudre le problème suivant

$$\inf_{x \in \mathbb{R}^n} \lambda_1(A(x)) \tag{1}$$

où $\lambda_1(X)$ est la plus grande valeur propre de $X = A(x)$, élément de l'espace des matrices symétriques d'ordre n noté $\mathcal{S}_n$ et où l'opérateur A défini par

$$\mathbb{R}^n \ni x \longmapsto A(x) := A_0 + \mathcal{A}x$$

est affine avec $A_0 \in \mathcal{S}_n$ et $\mathcal{A}$ un opérateur linéaire de $\mathbb{R}^n$ dans $\mathcal{S}_n$.
Nous cherchons à obtenir un algorithme du second ordre qui converge. En utilisant la théorie du $\mathcal{U}$-Lagrangien, nous présentons l'algorithme souhaité qui, sous certaines conditions, converge de manière quadratique. Nous commençons par une analyse du premier ordre de notre fonction $\lambda_1(X)$. En rappelant des résultats d'analyse convexe (chapitre 1), nous obtenons des propriétés pour cette fonction au premier ordre (chapitre 2) qui nous permettront de construire un algorithme dit *des valeurs propres approchées* (chapitre 3). Ensuite, afin d'améliorer la précision de notre algorithme, nous poursuivons notre analyse pour atteindre le second ordre. Grâce à la théorie du $\mathcal{U}$-Lagrangien (chapitre 4), nous développons la fonction $\lambda_1(X)$ au second ordre (chapitre 5) pour aboutir à l'algorithme recherché (chapitre 6). Enfin, en utilisant la dualité, nous établissons un rapport entre notre méthode faisceau du second ordre et les nouvelles *méthodes faisceaux de type proximal* (chapitre 7).

Motivations

Lorsque nous avons choisi ce sujet de mémoire, nous savions que c'était un domaine de recherche assez spécifique. Pour cette raison, nous voulions en savoir

davantage. L'article sur lequel nous nous sommes basées étant actuel, il s'appuyait sur de nouvelles théories. C'est pourquoi beaucoup de notions nous étaient peu familières, que ce soit en raison de nos connaissances ou de l'état d'avancement des recherches. Nous avons donc appris de nouveaux concepts. Ce mémoire nous a aidé à mieux nous organiser face à des théories inconnues et de synthétiser et retirer les idées globales d'un livre ou d'un article. Etant donné le temps limité, nous avons dû renoncer à pousser notre recherche en profondeur pour certains résultats et concepts. Vu les liens entre les différents chapitres de cet article, il nous était impossible de nous consacrer séparemment à une tâche particulière. Cependant, nous avons pu tirer un certain profit à travailler à deux simultanément. En effet, souvent nous pouvions surmonter une difficulté en mettant en commun l'avis de l'une et l'autre.

Notations

$\mathbb{R}$ représente l'ensemble des réels.
$\mathbb{R}^n$ désigne l'ensemble des vecteurs à composantes réels,
x, y des éléments de $\mathbb{R}^n$.
$\langle x, y\rangle := x^T y$ est le produit scalaire entre x et y.
$\|x\| := \sqrt{\langle x, x\rangle}$ est la norme euclidienne du vecteur $x \in \mathbb{R}^n$.
$\mathcal{S}_n$ est l'ensemble des matrices symètriques d'ordre n,
X, Y des matrices dans $\mathcal{S}_n$.
$\mathrm{Tr}(X)$ est la trace de la matrice X.
$\langle X, Y\rangle := \mathrm{Tr}(XY)$ est le *produit scalaire de Frobenius* entre deux matrices X et $Y \in \mathcal{S}_n$.
$\|X\| := \sqrt{\langle X, X\rangle}$ est la *norme de Frobenius* de la matrice X.
$\lambda_i(X)$ désigne la $i$ème valeur propre associée à la matrice d'ordre n X. Nous ordonnons les valeurs propres de la manière suivante

$$\lambda_1(X) \geq \lambda_2(X) \geq \ldots \lambda_n(X) \ .$$

$E_1(X)$ représente l'espace propre associé à la plus grande valeur propre $\lambda_1(X)$.
s.d.p. est l'abréviation de semi-défini(e)-positif(ve). En langage mathématique, nous notons une matrice X s.d.p. par

$$X \succeq 0 \ .$$

d.p. est l'abréviation de défini(e) positif(ve). Une matrice X d.p. se note

$$X \succ 0 \ .$$

$\mathcal{S}_n^+$ est le cône des matrices symétriques et s.d.p. d'ordre n.
$\mathcal{C}_n$ est l'ensemble des matrices symétriques, s.d.p et dont la trace vaut un.
$\sigma_C(d) := \sup_{g \in C}\langle g, d\rangle$ est la fonction support de l'ensemble non-vide C de $\mathbb{R}^n$,

pour $d \in I\!R^n$.
$F_C(d) := \text{Argmax}_{c \in C}\langle c, d\rangle$ est la face exposée de l'ensemble non vide C de $I\!R^n$ par rapport à $d \in I\!R^n$.
$\partial f(x)$ est le sous–différentiel de la fonction convexe à valeur finie f en $x \in I\!R^n$.
$\partial_\varepsilon f(x)$ est le sous–différentiel approché de la fonction convexe f en $x \in I\!R^n$.
$f'(x;d) := \sigma_{\partial f(x)}(d)$ est la dérivée directionnelle de la fonction convexe f dans la direction d en x.
$f'_\varepsilon(x;d) := \sigma_{\partial_\varepsilon f(x)}(d)$ est la dérivée directionnelle approchée de la fonction f dans la direction d en x.
$\mathcal{E}^\perp$ désigne le sous–espace orthogonal au sous–espace $\mathcal{E}$.
$\mathcal{A}^*$ est l'opérateur adjoint de l'opérateur linéaire $\mathcal{A}$ défini par

$$\langle x, \mathcal{A}^* y\rangle \;=\; \langle \mathcal{A}x, y\rangle \;.$$

$\text{proj}_{\mathcal{E}} : I\!R^n \longrightarrow \mathcal{E}$ est l'opérateur de projection sur le sous-espace $\mathcal{E}$.
$\text{proj}^*_{\mathcal{E}} : \mathcal{E} \longrightarrow I\!R^n$ est l'injection canonique $\mathcal{E} \ni e \mapsto e \oplus 0 \in I\!R^n$.
ri C est l'intérieur relatif de l'ensemble non vide C.
aff C est l'enveloppe affine de l'ensemble non vide $C \subset I\!R^n$.
co C est l'envelope convexe de l'ensemble $C \subset I\!R^n$.
span C est le sous–espace engendré par l'ensemble non vide $C \subset I\!R^n$.
$X^\dagger$ est l'inverse généralisée de *Moore-Penrose* de la matrice X définie par

$$X^\dagger = \sum_{i \text{ tq } \lambda_i(X)\neq 0} \frac{1}{\lambda_i(X)} q_i q_i^T$$

si $\sum_{i=1}^n \lambda_i(X) q_i q_i^T$ est la décomposition spectrale de X.
Im $\mathcal{E}$ est le sous–espace image de l'opérateur $\mathcal{E}$,
rg $\mathcal{E}$ est la dimension du sous–espace Im $\mathcal{E}$.
ker $\mathcal{E}$ est le noyau de l'opérateur $\mathcal{E}$.
dim $\mathcal{E}$ est la dimension du sous–espace $\mathcal{E}$.
$\mathcal{E} \oplus \mathcal{F}$ est la somme directe des sous–espaces $\mathcal{E}$ et $\mathcal{F}$.

Première partie

Le premier ordre

Le but de cette partie est de construire l'algorithme des *valeurs propres approchées*. Nous commençons par rappeler des résultats généraux d'analyse convexe et d'algèbre linéaire (chapitre 1). Ensuite, nous appliquons ces concepts à la fonction $\lambda_1(X)$ pour obtenir des propriétés caractérisant son comportement au premier ordre (paragraphes 2.1 - 2.3). Des résultats similaires seront obtenus pour la fonction composée $f := \lambda_1 \circ A$ (paragraphe 2.4). Ces propriétés nous permettront de construire (3.3) l'algorithme recherché par le processus suivant : en un point $x_k \in I\!R^n$,

1. Définir une direction de descente d_k tel qu'il soit possible de diminuer la valeur de f lorsqu'en partant de x_k, on suit la direction d_k (paragraphe 3.1).

2. Le long de cette direction, chercher un pas t_k à l'aide d'une recherche linéaire (paragraphe 3.2).

3. Poser $x_{k+1} = x_k + t_k d_k$.

Chapitre 1

Préliminaires

1.1 Eléments d'analyse convexe

1.1.1 Fonctions convexes et concaves

1.1.1.1 Définitions élémentaires

Soit C une partie non vide de $\mathbb{R}^n$.

• Ensemble convexe :

Une partie non vide C de $\mathbb{R}^n$ est **convexe** si

$$\forall\, x\,,\, y \in C\,,\, \forall\, \alpha \in]0,1[\qquad \alpha x + (1-\alpha)y \in C\ .$$

• Domaine et épigraphe :

Soit une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$, on définit
le **domaine de** f par

$$\text{Domf } = \{x \in \mathbb{R}^n \mid f(x) < +\infty\}\ ,$$

et l' **épigraphe de** f par

$$\text{Epi f } = \{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha\}\ .$$

• Fonction propre, convexe :

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ est dite **propre** si

$$\text{Domf } \neq \emptyset\ .$$

Elle est dite **convexe** si Epif est une partie convexe de $\mathbb{R}^n \times \mathbb{R}$.

• Fonction convexe :

La fonction $f : C \longrightarrow \mathbb{R}$ est appelée **convexe** sur la partie convexe C si $\forall\, x, y \in C, \forall\, \alpha \in]0,1[$:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) .$$

• Fonction concave :

La fonction $f : C \longrightarrow \mathbb{R}$ est appelée **concave** sur la partie convexe C si $\forall\, x, y \in C, \forall\, \alpha \in]0,1[$:

$$f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y) .$$

1.1.1.2 Fonctions convexes particulières

• Fonction affine :

Soit $a \in \mathbb{R}^n$, $b \in \mathbb{R}$. La fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ définie par $f(x) = a^T x - b$ est appelée **fonction affine** sur $\mathbb{R}^n$. Elle est convexe et concave sur $\mathbb{R}^n$. Si $b = 0$, alors f est dite **linéaire**. Elle "passe" par $(0,0)$ i.e. $f(0) = 0$.

• Fonction convexe fermée :

Soit une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$, f est **fermée** si son épigraphe Epif est une partie fermé de $\mathbb{R}^n \times \mathbb{R}$.

• Fonction support :

La **fonction support** associée à un ensemble C est la fonction $\sigma_C : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ définie par

$$\sigma_C(x) = \sup_{c \in C} \langle x, c \rangle$$

où $\langle x, c \rangle$ est le produit scalaire entre x et c.

L'ensemble

$$F_C(x) \;=\; \{c \in C \mid \sigma_C(x) = \langle x, c \rangle\}$$

est appelé **face exposée de C associée à x.**

Remarque 1.1.1 *Si C est un ensemble borné, alors* Dom $\sigma_C = \mathbb{R}^n$.

• Fonction conjuguée

Soit une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$, on définit f^*, la **fonction conjuguée** ou la **transformée de Fenchel** de f par

$$f^* : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$$

$$f^*(x) = \sup_{y \in \mathbb{R}^n} \{\langle x, y \rangle - f(y)\} .$$

La fonction f^* est **convexe, propre et fermée.**

1.1.2 Intérieur relatif

Certains ensembles convexes non-vides de $\mathbb{R}^n$ possèdent un intérieur vide. Pour échapper à ce genre de situation, l'idée est de prendre en compte l'**intérieur relatif**, intérieur de l'ensemble considéré non pas dans l'espace tout entier mais dans un ensemble particulier (son *enveloppe affine*, définie ci-dessous).

1.1.2.1 Définitions

• Combinaison affine :

Une **combinaison affine** d'éléments $x_1, \ldots, x_k$ de $\mathbb{R}^n$ est un élément de la forme

$$\sum_{i=1}^{k} \alpha_i x_i \quad \text{avec} \quad \sum_{i=1}^{k} \alpha_i = 1 \text{ et } \alpha_i \in \mathbb{R}.$$

• Combinaison convexe :

Une **combinaison convexe** d'éléments $x_1, \ldots, x_k$ de $\mathbb{R}^n$ est un élément de la forme

$$\sum_{i=1}^{k} \alpha_i x_i \quad \text{avec} \quad \sum_{i=1}^{k} \alpha_i = 1 \text{ et } \alpha_i \in \mathbb{R}^+ \qquad \forall\, i\ .$$

• Sous–espace affine :

Un **sous–espace affine** est le translaté d'un sous–espace vectoriel.

• Enveloppe affine :

L'**enveloppe affine** d'un ensemble C, notée aff C, est l'intersection de tous

les sous–espaces affines contenant l'ensemble C.

• Enveloppe convexe :

L'**enveloppe convexe** d'un ensemble C est l'ensemble des combinaisons convexes d'éléments de C. On la note co C.

• L'intérieur relatif :

L'**intérieur relatif**, ri C, d'un ensemble convexe $C \subset \mathbb{R}^n$ est l'intérieur de C pour la topologie relative à l'enveloppe affine de C, i.e.,

$$x \in \text{ri } C \Leftrightarrow \quad x \in \text{aff } C \quad \text{et} \quad \exists\, \delta > 0 \quad \text{tq} \quad \text{aff } C \cap B(x, \delta) \subset C \ .$$

Remarque 1.1.2 *Soit C, un sous-ensemble convexe de $\mathbb{R}^n$. Alors*

$$\begin{aligned} &i) \quad C \neq \emptyset \ \Rightarrow \ \text{ri } C \neq \emptyset \ , \\ &ii) \quad \text{int } C \neq \emptyset \ \Rightarrow \ \text{ri } C = \text{int } C \ . \end{aligned}$$

1.1.2.2 Exemples

1. $C = \{x\}$.
   Par définition, nous avons comme enveloppe affine

   $$\text{aff } C = \{x\}$$

   et comme intérieur relatif

   $$\text{ri } C = \{x\},$$

   alors que l'intérieur de C est vide.
2. $C = [a, b]$ avec $a \neq b \in \mathbb{R}^n$.
   L'enveloppe affine est la droite affine engendrée par a et b. L'intérieur relatif correspond à l'intervalle ouvert $]a, b[$ et l'intérieur est vide.

1.1.3 Sous–différentiabilité

Le but de cette section est d'approximer f dans un voisinage de x par une fonction linéaire.

1.1.3.1 Différentiabilité

Soit une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ et soit $x \in \text{int (Dom } f)$.
La fonction f est **différentiable en** x si il existe une fonction $l_x : \mathbb{R}^n \longrightarrow \mathbb{R}$, linéaire telle que

$$f(x+h) = f(x) + l_x(h) + o(\|h\|).$$

Remarque 1.1.3 *La convexité d'une fonction f n'entraîne pas automatiquement sa différentiabilité. La fonction valeur absolue illustre très bien cette remarque. En effet, $f(x) = |x|$ n'est pas différentiable en $x = 0$.*

1.1.3.2 Dérivées directionnelles

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$, $x_0 \in \text{Dom } f$ et $d \in \mathbb{R}^n$.
La **dérivée directionnelle** de f en x_0 dans la direction d est définie par

$$f'(x;d) = \lim_{t \searrow 0} \frac{f(x_0 + td) - f(x_0)}{t}$$

si la limite existe. Quand f est convexe, la limite peut être remplacée par l'infimum.

1.1.3.3 Le sous–différentiel

Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, et $x \in \text{Dom} f$.
Le **sous-différentiel** de f en x, noté $\partial f(x)$, est la partie convexe de $\mathbb{R}^n$ dont la fonction support est $f'(x;.)$. Il s'écrit :

$$\partial f(x) = \{s \in \mathbb{R}^n \mid \langle s, d\rangle \leq f'(x;d) \ \forall\, d \in \mathbb{R}^n\} \ .$$

Et, nous avons (montré dans [8])

$$f'(x;d) = \sup_{s \in \partial f(x)} \langle s, d\rangle \ = \ \sigma_{\partial f(x)}(d) \ .$$

Le vecteur $s \in \partial f(x)$ est appelé **sous-gradient** de f en x et peut être caractérisé comme suit :

$$s \in \partial f(x) \quad \Leftrightarrow \quad \forall\, y \in \mathbb{R}^n \quad f(y) \geq f(x) + \langle s, y - x\rangle \ .$$

Remarque 1.1.4 *Quand f est différentiable, le sous–differentiel est réduit au gradient, et on a*

$$f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0)$$

pour une fonction convexe, c'est-à-dire, la tangente est sous la courbe.

Proposition 1.1.1 *Pour le sous–différentiel d'une fonction f convexe, propre et fermée, on a*

$$x \in \text{ri Dom } f \;\Rightarrow\; \partial f(x) \neq \emptyset \; .$$

□.

1.1.3.4 Sous–différentiel pour une fonction supremum

1. Définitions et propriétés.

   Soit $(f_j)_{j \in J}$ une collection de fonctions convexes de $\mathbb{R}^n$ dans $\mathbb{R}$ avec J un ensemble dénombrable d'indices.
   Nous considérons $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ définie par

   $$f(x) = \sup_{j \in J} f_j(x) \; .$$

   Si Dom $f \neq \emptyset$, alors f est convexe et propre.

   Définition 1.1.1

   En chaque élément x_0 du Dom f, *l'**ensemble des indices actifs**, noté $J(x_0)$, est défini par*

   $$J(x_0) = \{j \in J \mid f_j(x_0) = f(x_0)\} \; .$$

   Le résultat suivant nous donne une méthode pour construire le sous–différentiel d'une telle fonction.

   Proposition 1.1.2 *Si $J = \{1, \ldots, m\}$, alors*

   $$\partial f(x) = \text{co}\{\cup \partial f_j(x) \mid j \in J(x)\} \; .$$

   □.

Corollaire 1.1.1 *Soient $f_1, \ldots, f_m$, m fonctions convexes différentiables de $\mathbb{R}^n \longrightarrow \mathbb{R}$, alors*

$$f = \sup_{1 \leq j \leq m} f_j \text{ est convexe}$$

et

$$\forall\, x \in \mathbb{R}^n, \quad \partial f(x) = \text{co}\{\nabla f_j(x) \,|\, j \in J(x)\}\ .$$

□.

2. Exemples.

Prenons $f : \mathbb{R}^2 \to \mathbb{R}$ où $f(x) = \max\{f_1(x), f_2(x), f_3(x)\}$ telle que

$$f_1(x) = -x_1 - x_2,\ f_2(x) = -x_1 + x_2 \text{ et } f_3(x) = x_1.$$

Nous cherchons $\partial f(4,8)$.
Puisque $J((4,8)) = \{2,3\}$, nous avons $\nabla f_2(4,8) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ et $\nabla f_3(4,8) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Par conséquent,

$$\partial f(4,8) = \text{co}\{(-1,1),(1,0)\}.$$

1.1.3.5 Sous–différentiel et directions de descente

Définition 1.1.2

Une direction $d \in \mathbb{R}^n$ est appelée une ***direction de descente*** *en x pour f si*

$$\exists\, \delta > 0 \quad \forall\, t \in\,]0, \delta] \quad f(x + td) < f(x)$$

Puisque la fonction f est convexe, cette propriété est équivalente à

$$\exists\, t > 0 \quad tq \quad f(x + td) < f(x)\ .$$

Proposition 1.1.3 *d est une direction de descente en x pour f*

$$\begin{aligned} &\Leftrightarrow \quad f'(x;d) < 0 \\ &\Leftrightarrow \quad \langle s, d\rangle < 0 \quad \forall s \in \partial f(x). \end{aligned}$$

□.

Proposition 1.1.4 *On peut trouver une direction de descente en x pour f*

$$\Leftrightarrow$$
$$0 \notin \partial f(x)$$

□.

Interprétation géométrique

Une direction de descente correspond à la normale d'un hyperplan séparant strictement l'ensemble convexe fermé $\partial f(x)$ de $\{0\}$.
Soit $x \in \mathbb{R}^n$ tel que $0 \notin \partial f(x)$. La **direction de plus forte descente** en x pour f est une solution du problème

$$\min_{\|d\| \leq 1} f'(x;d) \quad \Leftrightarrow \min_{\|d\| \leq 1} \max_{s \in \partial f(x)} \langle s, d \rangle.$$

Théorème 1.1.1 *Soit $x \in \mathbb{R}^n$ tel que $0 \notin \partial f(x)$, alors*

1. *le vecteur de norme minimale de $\partial f(x)$ existe et est unique; il coïncide avec la projection orthogonale de 0 sur $\partial f(x)$.*
2. *la direction de plus forte pente en x pour f est le vecteur $\frac{-m}{\|m\|}$ où m est le vecteur de norme minimale dans $\partial f(x)$.* □.

Pour résoudre un tel problème, l'algorithme de plus forte descente de Cauchy existe et utilise le sous–différentiel mais ne converge pas. Pour cette raison, nous ne l'expliciterons pas. L'idée pour s'en sortir est de considérer le sous–différentiel approché $\partial_\varepsilon f(x)$ au lieu du sous-différentiel $\partial f(x)$.

1.1.4 Le sous–différentiel approché

1.1.4.1 Définitions, exemples et propriétés de base

Soit f une fonction convexe propre, fermée et un point $x \in \text{Dom } f$. Soit aussi $\varepsilon \geq 0$. Le **ε-sous–différentiel** de f en x est l'ensemble défini par

$$\partial_\varepsilon f(x) = \{s \in \mathbb{R}^n \ : \ f(y) \geq f(x) + \langle s, y - x \rangle - \varepsilon \quad \forall\, y \in \text{Dom } f\}\ .$$

Les éléments de cet ensemble sont appelés des **ε-sous–gradients** de f en x.

Remarque 1.1.5 *Si $\varepsilon > 0$, alors*

$$\partial_\varepsilon f(x) \neq \emptyset \quad \textit{quand} \quad x \in \text{Dom } f\ .$$

On a aussi

$$\begin{aligned} \partial_\varepsilon f(x) &\subset \partial_{\varepsilon'} f(x) \quad \textit{pour } \varepsilon \leq \varepsilon' , \\ \partial f(x) &= \partial_0 f(x) = \cap_{\varepsilon>0} \partial_\varepsilon f(x) . \end{aligned}$$

Proposition 1.1.5 *Soit une fonction f convexe, propre, fermée et $\varepsilon \geq 0$, alors*

1. *$\partial_\varepsilon f(x)$ est fermé et convexe .*
2. *$\partial_\varepsilon f(x)$ est non vide et borné $\Leftrightarrow$ $x \in \operatorname{int} \operatorname{Dom} f$.*

□.

Le ε-sous–différentiel peut être caractérisé via la fonction conjuguée (ou Transformée de Fenchel).

Proposition 1.1.6 *Soit $x \in \operatorname{Dom} f$ et $s \in I\!R^n$, alors*

$$\begin{aligned} s \in \partial_\varepsilon f(x) &\Leftrightarrow f^*(s) + f(x) - \langle s, x \rangle \leq \varepsilon \\ &\Leftrightarrow x \in \partial_\varepsilon f^*(s) . \end{aligned}$$

□.

Proposition 1.1.7

$$0 \in \partial_\varepsilon f(x) \quad \Leftrightarrow \quad \forall y \in I\!R^n \; f(x) \leq f(y) + \varepsilon .$$

□.

1.1.4.2 La dérivée directionnelle approchée

Soit une fonction f convexe, propre, fermée. Soient $x \in \operatorname{Dom} f$ et $\varepsilon > 0$. La **ε-dérivée directionnelle** de f en x, notée $f'_\varepsilon(x; d)$ est définie par

$$f'_\varepsilon(x; d) = \inf_{t \searrow 0} \frac{f(x + td) - f(x) + \varepsilon}{t} .$$

Elle s'écrit également en termes de fonction support

$$f'_\varepsilon(x; d) = \sigma_{\partial_\varepsilon f(x)}(d) = \sup_{s \in \partial_\varepsilon f(x)} \langle s, d \rangle .$$

Remarque 1.1.6 *En $\varepsilon = 0$, on a $f'_\varepsilon(x; d) = f'(x; d)$.*

1.1.4.3 Sous-différentiel et directions d' ε-descente

Définition 1.1.3

Soit une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, convexe. Une direction $d \in \mathbb{R}^n$ est dite ***direction d'ε-descente*** *en x pour f si*

$$\exists\, t > 0 \qquad f(x + td) \leq f(x) - \varepsilon \ .$$

De façon similaire au paragraphe 1.1.3.5, nous avons les résultats suivants.

Proposition 1.1.8 *Les propriétés suivantes sont équivalentes*

i) d *est une direction d'*$\varepsilon-$*descente en* x *pour* f ,
ii) $f'_\varepsilon(x; d) < 0$,
iii) $\langle s, d\rangle < 0 \quad \forall\, s \in \partial_\varepsilon f(x)$.

□.

Proposition 1.1.9

i) $0 \notin \partial_\varepsilon f(x) \Leftrightarrow \exists\, d \quad f'_\varepsilon(x; d) < 0$
ii) $0 \notin \partial_\varepsilon f(x) \Rightarrow d = -\mathrm{proj}_{\partial_\varepsilon f(x)}\, 0$
et d *est une direction d'*$\varepsilon-$*descente en* x *pour* f .

□.

1.1.4.4 Algorithme conceptuel

A partir de l'algorithme de plus forte pente évoqué dans le paragraphe 1.1.3.5, nous construisons un algorithme d'ε-descente, qui est convergent.

Algorithme 1.1.1

PAS 1 : *Choisir $x_1 \in \mathbb{R}^n$ et $\varepsilon > 0$. Poser $k = 1$*

PAS 2 : *Si $0 \in \partial_\varepsilon f(x_k)$, alors on s'arrête et x_k est la ε-solution du problème de minimisation. Sinon, calculer d_k une direction de ε-descente en x_k pour f (e.g. prendre l'opposé de la projection de 0 sur $\partial_\varepsilon f(x)$.)*

PAS 3 : *Effectuer une recherche linéaire pour trouver le pas t_k, tel que*

$$f(x_k + t_k d_k) < f(x_k) - \varepsilon \ .$$

PAS 4 : *Mettre à jour x et k ; poser $x_{k+1} = x_k + t_k d_k$ et $k = k + 1$.*
Retourner au PAS 2. □.

Théorème 1.1.2 THÉORÈME DE CONVERGENCE*(voir [20], ch.7, th.1)*
Dans l'algorithme 1.1.1,

soit $f(x_k)$ *tend vers* $-\infty$, *quand* k *tend vers* $+\infty$,
soit on s'arrête après un nombre fini d'itérations
et on a un ε*−minimum de* f.

□.

1.2 Rappels d'algèbre linéaire

1.2.1 Valeurs propres

Définition 1.2.1

Soit A une transformation linéaire sur l'espace vectoriel K^n. On dira que $\lambda \in K$ est une ***valeur propre*** *de A associée au vecteur propre non nul $x \in K^n \backslash \{0\}$ si*

$$Ax = \lambda x.$$

L'ensemble des combinaisons linéaires des vecteurs propres associés à la valeur propre λ forme un sous–espace vectoriel, appelé ***sous–espace propre*** *associé à la valeur propre λ.*

Définition 1.2.2

Soit A une transformation linéaire sur E. La ***transformation adjointe*** *de A par rapport au produit scalaire, notée A^*, est la transformation linéaire sur E telle que*

$$\langle Ax, y \rangle = \langle x, A^* y \rangle$$

pour tout $x, y \in E$.

Définition 1.2.3

Une transformation linéaire A d'un espace métrique E dans lui-même est dite ***auto-adjointe*** *si et seulement si $A = A^*$.*

Proposition 1.2.1 *Soit A la matrice d'une transformation linéaire auto-adjointe d'un espace métrique dans lui-même, par rapport à une base orthonormale X fixée. Alors on a la relation $[A]_{ij} = \overline{[A]}_{ji}$.* □.

Une matrice satisfaisant cette propriété est appelée **symétrique** quand l'espace métrique sous-jacent est $\mathbb{R}^n$ et **hermitienne** lorsque cet espace est $\mathbb{C}^n$.

Définition 1.2.4

*Soit A une matrice hermitienne. Le **quotient de Rayleigh** qui lui est associé est la fonction*

$$r_A(x) = \frac{\langle x, Ax\rangle}{\langle x, x\rangle} \qquad \forall\, x \neq 0\ .$$

Définition 1.2.5

*Une matrice hermitienne A est dite **semi définie positive** si toutes ses valeurs propres sont positives.*

1.2.2 Trace d'une matrice symétrique

Définition 1.2.6

*La **trace** d'une matrice X carrée d'ordre n est définie par*

$$\mathrm{Tr}(X) = \sum_{i=1}^{n} X_{ii} \qquad \text{où } X_{ii} \text{ sont les éléments diagonaux de } X$$

ou encore

$$\mathrm{Tr}(X) = \sum_{i=1}^{n} \lambda_i(X) \qquad \text{où } \lambda_i(X) \text{ sont les valeurs propres de } X\ .$$

Définition 1.2.7

*On définit le **produit scalaire de Frobenius** de deux matrices carrées X, Y d'ordre n par*

$$\langle X, Y\rangle = \mathrm{Tr}(XY).$$

Définition 1.2.8

*On définit la **norme de Frobenius** d'une matrice symétrique X d'ordre n comme étant*

$$\|X\| = \sqrt{\langle X, X\rangle}\ .$$

Proposition 1.2.2 *Soient deux matrices carrées A et B d'ordre n et $\alpha \in \mathbb{R}$. Alors*

1. $\mathrm{Tr}(\alpha A) = \alpha \mathrm{Tr}(A)$

2. $\mathrm{Tr}(A+B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$
3. $\mathrm{Tr}(AB) = \mathrm{Tr}(BA) \neq \mathrm{Tr}(A) \cdot \mathrm{Tr}(B)$
4. $\mathrm{Tr}(A) = \mathrm{Tr}(HAH^{-1})$ *pour toute matrice* H *non singulière*
5. $\mathrm{Tr}(A) = \mathrm{Tr}(UAU^T)$ *pour toute matrice* U *telle que* $U^TU = I$
6. $\langle PAP^T, B\rangle = \langle A, P^TBP\rangle$ *pour toute matrice* P *d'ordre* n. □.

Proposition 1.2.3 ***Conséquence du théorème du produit de Schur :*** *Si* A *et* B *sont deux matrices carrées s.d.p., alors*

$$\mathrm{Tr}(AB) = 0 \Leftrightarrow AB = 0.$$

□.

1.2.3 Propriétés particulières de la trace faisant intervenir des matrices diagonales

Lemme 1.2.1 *Soient* D *une matrice diagonale de* $\mathbb{R}^{n\times n}$, U *une matrice orthogonale de* $\mathbb{R}^{n\times n}$ *telle que* $U^TU = I$ *et* X *une matrice symétrique de* $\mathbb{R}^{n\times n}$, *alors on a*

$$\langle D, U^TXU\rangle = \sum_{i=1}^{n} D_{ii}u_i^TXu_i$$

où les u_i *sont les colonnes de la matrice* U *et* D_{ii} *les éléments diagonaux de la matrice* D.

Preuve (dans le cas $n = 2$). Notons u_{ij}, la $j^{\text{ème}}$ composante de la $i^{\text{ème}}$ colonne de U. Alors

$$\begin{aligned}
U^TXU &= \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}\begin{bmatrix} a & c \\ c & b \end{bmatrix}\begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix} \\
&= \begin{bmatrix} au_{11}+cu_{12} & cu_{11}+bu_{12} \\ au_{21}+cu_{22} & cu_{21}+bu_{22} \end{bmatrix}\begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix} \\
&= \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \\
\text{où}\quad r_{11} &= au_{11}u_{11}+cu_{12}u_{11}+cu_{11}u_{12}+bu_{12}u_{12}, \\
r_{12} &= au_{11}u_{21}+cu_{12}u_{21}+cu_{11}u_{22}+bu_{12}u_{22}, \\
r_{21} &= au_{21}u_{11}+cu_{22}u_{11}+cu_{21}u_{12}+bu_{22}u_{12}, \\
r_{22} &= au_{21}u_{21}+cu_{22}u_{21}+cu_{21}u_{22}+bu_{22}u_{22}.
\end{aligned}$$

De plus, D étant diagonale,

$$D = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}.$$

Donc,

$$DU^T XU = \begin{bmatrix} d_1 r_{11} & d_1 r_{12} \\ d_2 r_{21} & d_2 r_{22} \end{bmatrix}.$$

Comme

$$r_{ii} = u_i^T X u_i,$$

nous avons

$$\mathrm{Tr}(DU^T XU) = d_1 r_{11} + d_2 r_{22} = \sum_{i=1}^{2} D_{ii} u_i^T X u_i.$$

□.

Lemme 1.2.2 *Soient D une matrice diagonale de $\mathbb{R}^{n\times n}$, U et V deux matrices orthogonales de $\mathbb{R}^{n\times n}$ et X une matrice symétrique de $\mathbb{R}^{n\times n}$, on a*

$$\langle D, U^T XV + V^T XU \rangle = \sum_{i=1}^{n} D_{ii} \langle X, u_i v_i^T + v_i u_i^T \rangle$$

où u_i et v_i sont respectivement les colonnes de la matrice U et V, et D_{ii} est l'élément diagonal de la $i^{ème}$ ligne de la matrice D.

Preuve (dans le cas n=2). Par définition de la trace, nous avons

$$\begin{aligned} \langle D, U^T XV + V^T XU \rangle &= \langle D, U^T XV \rangle + \langle D, V^T XU \rangle \\ &= \sum_{i=1}^{n} D_{ii} (U^T XV)_{ii} + \sum_{i=1}^{n} D_{ii} (V^T XU)_{ii}. \end{aligned}$$

Examinons d'abord l'élément diagonal de $U^T XV$. Nous avons,

$$\begin{aligned} \begin{bmatrix} u_1^T \\ u_2^T \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix} &= \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{bmatrix} \\ &= \begin{bmatrix} u_{11}a + u_{12}b & u_{11}b + u_{12}c \\ u_{21}a + u_{22}b & u_{21}b + u_{22}c \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{bmatrix} \\ &= \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \end{aligned}$$

où

$$r_{11} = v_{11}u_{11}a + v_{11}u_{12}b + v_{12}u_{11}b + v_{12}u_{12}c$$

$$r_{22} = u_{21}av_{21} + u_{22}bv_{21} + u_{21}bv_{22} + u_{22}cv_{22}.$$

Examinons d'autre part la trace de $(Du_iv_i^T)$, avec $i = 1$.

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} u_{11}v_{11} & u_{11}v_{12} \\ u_{12}v_{11} & u_{12}v_{12} \end{bmatrix}$$
$$= \begin{bmatrix} au_{11}v_{11} + bu_{12}v_{11} & au_{11}v_{12} + bv_{12}u_{12} \\ bu_{11}v_{11} + cu_{12}v_{11} & bu_{11}v_{12} + cu_{12}v_{12} \end{bmatrix}.$$

La trace vaut donc

$$au_{11}v_{11} + bu_{12}v_{11} + bu_{11}v_{12} + cu_{12}v_{12},$$

ce qui correspond bien à l'élément diagonal de U^TXV. On raisonne de manière analogue pour l'élément diagonal de V^TXU. En regroupant les termes, nous arrivons bien à la thèse. □.

Lemme 1.2.3 *Soient X une matrice symétrique de $\mathbb{R}^{r\times r}$ et U une matrice orthogonale de $\mathbb{R}^{n\times r}$, on a*

$$\text{Tr } (UXU^T) = \text{Tr } (X).$$

Preuve (dans le cas où n=3 et r=2). Examinons d'abord les éléments diagonaux de (UXU^T). Nous avons

$$UXU^T = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \end{bmatrix}$$
$$= \begin{bmatrix} au_{11} + bu_{12} & bu_{11} + cu_{12} \\ au_{21} + bu_{22} & bu_{21} + cu_{22} \\ au_{31} + bu_{32} & bu_{31} + cu_{32} \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \end{bmatrix}$$
$$= \begin{bmatrix} d_1 & (*) & (*) \\ (*) & d_2 & (*) \\ (*) & (*) & d_3 \end{bmatrix}.$$

Nous ne nous intéressons qu'aux éléments diagonaux qui sont

$$\begin{aligned} d_1 &= au_{11}^2 + bu_{11}u_{12} + bu_{12}u_{12} + cu_{12}^2, \\ d_2 &= au_{21}^2 + bu_{22}u_{21} + bu_{21}u_{22} + cu_{22}^2, \\ d_3 &= au_{31}^2 + bu_{31}u_{32} + bu_{31}u_{32} + cu_{32}^2. \end{aligned}$$

Etant donné que $U^TU = I$, en faisant la somme des éléments diagonaux, il ne reste dans la somme que les éléments diagonaux de la matrice X. En effet, en regroupant les termes dépendant des différents éléments de la matrice X, nous obtenons

$$\begin{aligned} \text{Tr } (UXU^T) &= a\,[u_{11}^2 + u_{21}^2 + u_{31}^2] \\ &\quad +b\,[u_{11}u_{12} + u_{12}u_{12} + u_{22}u_{21} + u_{21}u_{22} + u_{31}u_{32} + u_{31}u_{32}] \\ &\quad +c\,[u_{12}^2 + u_{22}^2 + u_{32}^2] \\ &= a + c\,. \end{aligned}$$

□.

1.2.4 Normes matricielles

Définition 1.2.9

Soit $\|\cdot\|$ une application de $\mathbb{C}^{n\times n}$ dans $\mathbb{R}$. Cette application est une ***norme matricielle*** *si les conditions suivantes sont satisfaites pour deux matrices carrées A et B de dimension n.*

- $\|A\| \geq 0$
- $\|A\| = 0 \Leftrightarrow A = 0$
- $\|hA\| = |h| \cdot \|A\|$ *pour tout nombre complexe h*
- $\|A + B\| \leq \|A\| + \|B\|$
- $\|AB\| \leq \|A\| \cdot \|B\|$.

Définition 1.2.10

Les ***valeurs singulières*** *d'une matrice (non nécessairement carrée) A sont les racines carrées des valeurs propres de A^*A.*

Remarque 1.2.1 *Toutes les normes matricielles sont équivalentes.*

1.2.5 Diagonalisation d'une matrice

Définition 1.2.11

Une matrice carrée A est ***diagonalisable*** *si elle est semblable à une matrice diagonale D, i.e. il existe une matrice régulière P telle que*

$$A = P^{-1}DP \ .$$

Dans le cas où D est la matrice des valeurs propres , on peut dire que

$$A = V^{-1}DV,$$

où V est la matrice des vecteurs propres orthonormés. $V^{-1}DV$ est la décomposition spectrale de la matrice A.

Remarque 1.2.2 *Toute matrice symétrique réelle est diagonalisable dans $\mathbb{R}$.*

Chapitre 2

Analyse du premier ordre

Le but de ce chapitre est de trouver une "bonne" direction de descente. A cette fin, nous allons analyser les différentes caractéristiques des sous–différentiel et sous–différentiel approché de la fonction "valeur propre maximale" λ_1 . Nous en déduirons des résultats similaires pour la fonction composée $f = \lambda_1 \circ A$ de $\mathbb{R}^n$ dans $\mathbb{R}$ qui, à un vecteur x, fait correspondre $\lambda_1(A(x))$, grâce à un simple développement en chaîne. Enfin, cette étude nous mènera au résultat principal de notre chapitre : toute direction d séparant 0 de l'élargissement choisi de $\partial f(x)$ est une "bonne" direction de descente .

2.1 Sous–différentiel et sous–différentiel approché de la fonction λ_1

Dans cette analyse, nous serons souvent confrontés à un ensemble particulier : l'intersection du cône des matrices semi-définies avec l'hyperplan des matrices symétriques dont la trace est égale à un . On le note $\mathcal{C}_n$ et il est défini par

$$\mathcal{C}_n = \{V \in \mathcal{S}_n : V \succeq 0,\ \mathrm{Tr}V = 1\}. \tag{2.1}$$

De plus $\mathcal{C}_n$ est un ensemble convexe et compact (cfr. **Annexe III.1**). Le lemme suivant, que l'on démontre via la décomposition spectrale des matrices symétriques, donne une description de l'ensemble $\mathcal{C}_n$.

Lemme 2.1.1 *L'ensemble convexe $\mathcal{C}_n$ est caractérisé par*

$$\mathcal{C}_n = \mathrm{co}\{qq^T : q \in \mathbb{R}^n,\ \|q\| = 1\} .$$

Preuve. (cfr. **Annexe III.2**) □.

En utilisant *la formulation variationnelle du quotient de Rayleigh* (définition 1.2.4).

$$\lambda_1(X) = \max_{q \in \mathbb{R}^n, \|q\|=1} q^T X q$$

et le lemme 2.1.1, nous obtenons une formulation en termes de fonction support pour λ_1 :

$$\lambda_1(X) = \sigma_{\mathcal{C}_n}(X) \tag{2.2}$$

La formulation de $\lambda_1(X)$ en terme de fonction support sera notre outil principal dans l'analyse du développement vertical de $\lambda_1(X)$. Nous avons également besoin de la description des faces exposées de $\mathcal{C}_n$.

$$F_{\mathcal{C}_n}(X) = \{Z \in \mathcal{C}_n \text{ telle que } \langle X, Z\rangle = \sigma_{\mathcal{C}_n}(X)\} \ . \tag{2.3}$$

Le résultat technique suivant démontré en annexe (cfr. **Annexe III.3**), nous permettra de donner une nouvelle expression pour les faces exposées de $\mathcal{C}_n$.

Lemme 2.1.2 *Soient X et $Z \in \mathcal{S}_n$, et $Q = [q_1 \dots q_r]$ une matrice $n \times r$ dont les colonnes forment une base orthonormale de* $\ker X$.

(*i*) *Alors $XZ = 0$ si et seulement si*
$\exists\ Y \in \mathcal{S}_r$ *telle que* $Z = QYQ^T$.

(*ii*) *Supposons en outre que X , $Z \in \mathcal{S}_n^+$,*
alors $\langle X, Z\rangle = 0$ si et seulement si
$\exists\ Y \succeq 0$ *telle que* $Z = QYQ^T$.

□.

Dès lors, nous pouvons énoncer le théorème :

Théorème 2.1.1

(i) Soit $X \in \mathcal{S}_n$ et Q_1 une matrice $n \times r$ dont les colonnes forment une base orthonormale de $E_1(X)$. La face de $\mathcal{C}_n$ exposée par rapport à X est

$$F_{\mathcal{C}_n}(X) = \{Q_1 Y Q_1^T |\ Y \in \mathcal{C}_r\} = \text{co}\{qq^T \mid \|q\| = 1, q \in E_1(X)\} \ . \tag{2.4}$$

(ii) Le sous-différentiel de λ_1 en X est

$$\partial\lambda_1(X) = F_{\mathcal{C}_n}(X) \ . \tag{2.5}$$

Preuve.

(i) Tout d'abord, nous avons que

$$\{Q_1 Y Q_1^T |\ Y \in \mathcal{C}_r\} = \text{co}\{qq^T :\ \|q\| = 1, q \in E_1(X)\} \ . \tag{2.6}$$

En effet, nous pouvons écrire tout vecteur normalisé de $E_1(X)$ sous la forme $q = Q_1 z$, avec $z \in I\!R^r$ et $\|z\| = 1$ (car les colonnes de Q_1 forment une base de $E_1(X)$).
Ainsi, nous avons par le lemme 2.1.1,

$$\begin{aligned} \mathrm{co}\{qq^T : \|q\| = 1,\, q \in E_1(X)\} &= \mathrm{co}\{Q_1 z z^T Q_1^T : \|z\| = 1 \text{ et } z \in I\!R^r\} \\ &= Q_1\, \mathrm{co}\{zz^T : \|z\| = 1 \text{ et } z \in I\!R^r\}\, Q_1^T \\ &= Q_1 \mathcal{C}_r Q_1^T \\ &= \{Q_1 Y Q_1^T : Y \in \mathcal{C}_r\}. \end{aligned}$$

Montrons ensuite que la première égalité de (2.4) est satisfaite. Par la définition d'une face exposée (2.3) et les égalités (2.1) et (2.2), nous avons immédiatement les équivalences suivantes :

$$\begin{aligned} Z \in F_{\mathcal{C}_n}(X) &\Leftrightarrow Z \in \mathcal{C}_n \text{ et } \langle X, Z\rangle = \sigma_{\mathcal{C}_n}(X) \\ &\Leftrightarrow Z \in \mathcal{C}_n \text{ et } \langle \lambda_1(X) I_n - X, Z\rangle = 0\ . \end{aligned} \tag{2.7}$$

Pour montrer que (2.7) est équivalent à dire qu'il existe une matrice $Y \in \mathcal{C}_r$ tel que $Z = Q_1 Y Q_1^T$, nous allons utiliser le lemme 2.1.2 (ii). Pour cela, remarquons que

$$\lambda_1(X) I_n - X \in \mathcal{S}_n^+ \text{ et } Z \in \mathcal{S}_n^+.$$

En effet, X étant une matrice symétrique et les valeurs propres de $\lambda_1(X) I_n - X$ étant toutes positives puisque $\lambda_1(X)$ est la plus grande valeur propre de X, la première condition est satisfaite. Et d'autre part, la matrice Z étant dans l'ensemble $\mathcal{C}_n$, est aussi dans $\mathcal{S}_n^+$. Dès lors, par le lemme 2.1.2,

$$\langle \lambda_1(X) I_n - X, Z\rangle = 0$$

est équivalent à

$$\exists\, Y \succeq 0 \quad \text{tq} \quad Z = QYQ^T$$

où $Q = [q_1 \dots q_r]$ est une matrice dont les colonnes forment une base orthonormale de $\ker(\lambda_1(X) I_n - X)$. Comme $\ker(\lambda_1(X) I_n - X) = \{u \in I\!R^n \mid \lambda_1(X) u = Xu\} = E_1(X)$ et que $\mathrm{Tr} Z = 1$, nous obtenons l'équivalence annoncée.

(ii) Puisque $\partial \sigma_{\mathcal{C}_n}(X) = F_{\mathcal{C}_n}(X)$ (voir par exemple dans [8], chapitre 6, exemple 3.1, p.258), avec l'égalité (2.2), nous pouvons écrire

$$\partial \lambda_1(X) = \partial \sigma_{\mathcal{C}_n}(X) = F_{\mathcal{C}_n}(X)\ ,$$

c'est-à-dire (2.5). □.

La description du sous–différentiel approché est aussi obtenue directement à partir de la formulation de $\lambda_1(X)$ en terme de fonction support :

Théorème 2.1.2 *$\forall \varepsilon \geq 0$, on a*

$$\partial_\varepsilon \lambda_1(X) = \{Z \in \mathcal{C}_n : \langle X, Z\rangle \geq \lambda_1(X) - \varepsilon\} . \tag{2.8}$$

Preuve. Nous avons successivement

$$\begin{aligned} \partial_\varepsilon \lambda_1(X) &= \partial_\varepsilon \sigma_{\mathcal{C}_n}(X) \text{ par (2.2)} \\ &= \{Z \in \mathcal{C}_n : \sigma_{\mathcal{C}_n}(X) \leq \langle Z, X\rangle + \varepsilon\} \text{ par l'exemple 1.2.5 de [8]} \\ &= \{Z \in \mathcal{C}_n : \langle Z, X\rangle \geq \lambda_1(X) - \varepsilon\} . \end{aligned}$$

Nous obtenons ainsi (2.8) . □.

Le théorème 2.1.1 nous permet d'exprimer facilement la dérivée directionnelle de λ_1.

Théorème 2.1.3 *Soient X, $D \in \mathcal{S}_n$ et Q_1 une matrice dont les colonnes forment une base orthonormale de $E_1(X)$. Alors*

$$\lambda_1'(X; D) = \lambda_1(Q_1^T D Q_1) . \tag{2.9}$$

Preuve. La dérivée directionnelle en x d'une fonction convexe étant égale à la fonction support de son sous–différentiel en x, nous obtenons, en utilisant la définition de fonction support, (2.4), (2.5) et enfin l'égalité (2.1), les égalités suivantes

$$\begin{aligned} \lambda_1'(X; D) &= \sigma_{\partial\lambda_1(X)}(D) \\ &= \max_{Z \in F_{\mathcal{C}_n}} \langle D, Z\rangle \\ &= \max_{Y \in \mathcal{C}_r} \langle D, Q_1 Y Q_1^T\rangle \\ &= \max_{Y \in \mathcal{C}_r} \langle Q_1^T D Q_1, Y\rangle \\ &= \lambda_1(Q_1^T D Q_1) . \end{aligned}$$

□.

En optimisation non-différentiable, il est bien connu que la propriété de descente $\lambda_1'(X, D) < 0$ d'une direction de recherche D ne conduit pas à des algorithmes convergents. Comme expliqué en [8], les algorithmes de minimisation basés sur cette propriété peuvent être numériquement inefficaces. Pour éviter ce problème, nous allons nous baser sur la philosophie suivie dans [8], à savoir considérer les directions d'ε-descente. Une direction D est une direction d'ε-descente si la ε-dérivée directionnelle en X définie par

$$\lambda_{1,\varepsilon}'(X; D) = \sigma_{\partial_\varepsilon \lambda_1(X)}(D)$$

est strictement négative.
Géométriquement, ce sont les directions qui séparent 0 de $\partial_\varepsilon \lambda_1(X)$. La fonction $X \longmapsto \lambda_{1,\varepsilon}'(X; D)$ étant continue ([8], th.4.1.3 du ch.11), l'efficacité numérique de telles directions est à présent garantie. L'algorithme de séparation exposé en ([8], ch.13 et ch.14) fournit de telles directions. Comme $\partial_\varepsilon \lambda_1(X)$ n'est pas explicitement connu, notre stratégie consiste à séparer 0, non pas de $\partial_\varepsilon \lambda_1(X)$, mais d'une assez "bonne approximation" de cet ensemble.

2.2 Elargissement du sous–différentiel

En vue de définir une bonne approximation de $\partial_\varepsilon \lambda_1(X)$, introduisons les concepts suivants :

Définition 2.2.1

Pour toute matrice $X \in \mathcal{S}_n$ et $\varepsilon \geq 0$, on définit :

- *l'ensemble des indices des ε-plus grandes valeurs propres*

$$I_\varepsilon(X) := \{i \in \{1, \ldots, n\} \, : \, \lambda_i(X) \geq \lambda_1(X) - \varepsilon\}, \tag{2.10}$$

- *l'ε-multiplicité de $\lambda_1(X)$:*

$$r_\varepsilon := \max\{i \, : \, i \in I_\varepsilon(X)\}, \tag{2.11}$$

- *le ε-premier espace propre :*

$$E_\varepsilon(X) := \oplus_{i \in I_\varepsilon(X)} E_i(X), \tag{2.12}$$

  où $E_i(X)$ est l'espace propre associé à la $i^{\text{ème}}$ valeur propre $\lambda_i(X)$,

- *son complément orthogonal :*

$$F_\varepsilon(X) := \oplus_{i \notin I_\varepsilon(X)} E_i(X), \tag{2.13}$$

– *la "séparation spectrale" à ε près :*

$$\Delta_\varepsilon(X) = \lambda_{r_\varepsilon}(X) - \lambda_{r_\varepsilon+1}(X). \tag{2.14}$$

Pour chaque matrice $X \in \mathcal{S}_n$ et $\varepsilon \geq 0$, nous pouvons définir l'ensemble suivant :

$$\delta_\varepsilon\lambda_1(X) = \text{co}\{ee^T : \|e\| = 1\,,\ e \in E_\varepsilon(X)\}\,. \tag{2.15}$$

Cet ensemble est convexe et compact (cfr. **Annexe III.4**). Si nous désignons par Q_ε une matrice $n \times r_\varepsilon$ dont les colonnes forment une base de $E_\varepsilon(X)$, il suit du théorème 2.1.1 et de l'**Annexe III.5** que

$$\delta_\varepsilon\lambda_1(X) = \{Q_\varepsilon Y Q_\varepsilon^T : Y \in \mathcal{C}_{r_\varepsilon}\} = F_{\mathcal{C}_n}(Q_\varepsilon Q_\varepsilon^T) = \partial\lambda_1(Q_\varepsilon Q_\varepsilon^T)\,. \tag{2.16}$$

Cet ensemble est une approximation externe de $\partial\lambda_1(X)$ et une approximation interne de $\delta_\varepsilon\lambda_1(X)$:

Proposition 2.2.1 *Soit $X \in \mathcal{S}_n$. Alors, $\forall\,\varepsilon \geq 0$, on a :*

$$\partial\lambda_1(X) \subset \delta_\varepsilon\lambda_1(X) \subset \partial_\varepsilon\lambda_1(X). \tag{2.17}$$

Preuve. Montrons la première inclusion : $\partial\lambda_1(X) \subset \delta_\varepsilon\lambda_1(X)$. Nous savons que

$$\begin{aligned}\partial\lambda_1(X) &= F_{\mathcal{C}_n}(X)\\ &= \text{co}\{qq^T : \|q\| = 1\,,\ q \in E_1(X)\end{aligned}$$

et

$$\delta_\varepsilon\lambda_1(X) = \text{co}\{ee^T\,,\ \|e\| = 1\,,\ e \in E_\varepsilon(X)\}.$$

Or par (2.12),

$$E_1(X) \subset E_\varepsilon(X).$$

Dès lors

$$\partial\lambda_1(X) \subset \delta_\varepsilon\lambda_1(X).$$

Montrons la deuxième inclusion : $\delta_\varepsilon\lambda_1(X) \subset \partial_\varepsilon\lambda_1(X)$.

Nous procédons en deux étapes :

1. $\delta_\varepsilon\lambda_1(X) \subset \mathcal{C}_n$.
   En effet, par définition de l'élargissement (2.15),
   $$\delta_\varepsilon\lambda_1(X) = \text{co}\{ee^T : \|e\| = 1,\ e \in E_\varepsilon(X)\}$$
   et par le lemme 2.1.1,
   $$\mathcal{C}_n = \text{co}\{qq^T : q \in I\!R^n,\ \|q\| = 1\}.$$
   Comme $E_\varepsilon(X) \subset I\!R^n$, nous avons l'inclusion souhaitée.

2. $\delta_\varepsilon\lambda_1(X) \subset \partial_\varepsilon\lambda_1(X)$.
   Soit $Z \in \delta_\varepsilon\lambda_1(X)$. Grâce à (2.16),
   $$Z = Q_\varepsilon Y Q_\varepsilon^T \text{ avec } Y \in \mathcal{C}_{r_\varepsilon}\ .$$
   Par conséquent,
   $$\langle Z, X\rangle \ = \ \langle Q_\varepsilon Y Q_\varepsilon^T, X\rangle \ = \ \langle Y, Q_\varepsilon^T X Q_\varepsilon\rangle\ .$$
   Or,
   $$Q_\varepsilon^T X Q_\varepsilon = \ \text{diag}(\lambda_1(X), \ldots, \lambda_{r_\varepsilon}(X)) \succeq \lambda_{r_\varepsilon} I_{r_\varepsilon},$$
   d'où
   $$\langle Y, Q_\varepsilon^T X Q_\varepsilon\rangle \ \geq \ \langle Y, \lambda_{r_\varepsilon} I_{r_\varepsilon}\rangle.$$
   Et comme $\langle Y, I_{r_\varepsilon}\rangle = 1$ (car $Y \in \mathcal{C}_{r_\varepsilon}$),
   $$\langle Z, X\rangle \ \geq \ \lambda_{r_\varepsilon}.$$
   De plus, par la définition de r_ε (2.11),
   $$\lambda_{r_\varepsilon} \ \geq \ \lambda_1(X) - \varepsilon$$
   donc,
   $$\langle Z, X\rangle \ \geq \ \lambda_1(X) - \varepsilon\ .$$
   Avec le fait que
   $$Z \in \delta_\varepsilon\lambda_1(X) \subset \mathcal{C}_n,$$
   on peut déduire du théorème 2.1.2 que, $Z \in \partial_\varepsilon\lambda_1(X)$. □.

Un point crucial consiste à quantifier la distance entre notre élargissement $\delta_\varepsilon\lambda_1(X)$ et le sous–différentiel approché $\partial_\varepsilon\lambda_1(X)$. Une façon de procéder est d'obtenir un développement vertical de λ_1.

2.3 Développement vertical

2.3.1 Présentation du développement vertical

Pour étudier le comportement du second ordre d'une fonction convexe à valeurs réelles, deux approches sont proposées. La première est appelée *"développement horizontal"* et consiste à étudier la limite du rapport

$$H(x,h) = \frac{f(x+h) - f(x) - f'(x,h)}{\frac{1}{2}\|h\|^2}$$

lorsque $\|h\| \to 0$.
La seconde, appelée *"développement vertical"* étudie la limite du rapport

$$V(x,d,\varepsilon) = \frac{[f'_\varepsilon(x;d) - f'(x,d)]^2}{2\varepsilon}$$

lorsque $\varepsilon \to 0$.
Dans un repère orthonormé, nous situons les abscisses sur l'axe horizontal et les images sur l'axe vertical. Dans le premier cas, h porte le nom d'incrément horizontal ; en effet h modifie les abscisses de f et varie sur l'axe horizontal. L'approche verticale doit son nom à la situation de l'incrément vertical ε sur l'axe vertical. En effet, en reprenant la définition de la dérivée directionnelle approchée pour une fonction convexe $f'_\varepsilon(x;d)$, le paramètre ε est sur le même pied que les images. Et donc dans notre repère, il se situe sur l'axe des ordonnées (axe vertical).

L'article [15] donne d'autres informations sur les conditions de croissance inférieure et supérieure et tire des conclusions sur la dérivée directionnelle de λ_1. Ces derniers résultats sont détaillés dans le paragraphe suivant.

2.3.2 Rappels spécifiques d'algèbre linéaire

Avant de présenter le théorème 2.3.1 qui donne une borne supérieure pour la distance entre le sous–différentiel approché et l'élargissement du sous–différentiel, nous rappelons quelques résultats particuliers d'algèbre linéaire sur lesquels nous nous appuyerons pour démontrer le théorème.

Lemme 2.3.1 *Soit $X \in \mathcal{S}_n$. Soit $U \in \mathbb{R}^{n\times n}$ tel que $U^TU = I_n$. Alors, il existe des matrices $n \times n$ $(E_\varepsilon, F_\varepsilon, \Sigma, T)$ telles que*

$$\begin{cases} \text{les colonnes de } E_\varepsilon \text{ sont des vecteurs unité de } E_\varepsilon(X) & (a) \\ \text{les colonnes de } F_\varepsilon \text{ sont des vecteurs unité de } F_\varepsilon(X) & (b) \\ \Sigma \text{ et } T \text{ sont diagonales et s.d.p.} & (c) \\ \Sigma^2 + T^2 = I & (d) \\ U = E_\varepsilon\Sigma + F_\varepsilon T\ . & (e) \end{cases} \tag{2.18}$$

Preuve. (cfr. **Annexe III.6**) □.

Lemme 2.3.2 *Soit $X \in \mathcal{S}_n$. Soit $(E_\varepsilon, F_\varepsilon, \Sigma, T)$ satisfaisant (2.18a),(2.18b),(2.18c) et (2.18d). Soit $\Theta = \mathrm{diag}(\theta_1, \ldots, \theta_n) \in \mathcal{C}_n$. Alors, on a :*

$$\begin{cases} E_\varepsilon^T A F_\varepsilon = F_\varepsilon^T A E_\varepsilon = 0 & (a) \\ \lambda_{r_\varepsilon}(X) \leq \langle X, E_\varepsilon \Theta F_\varepsilon^T \rangle \leq \lambda_1(X) & (b) \\ \langle X, F_\varepsilon \Theta F_\varepsilon^T \rangle \leq \lambda_{r_\varepsilon+1}(X) & (c) \\ \mathrm{Tr}(\Sigma \Theta T) \leq [\mathrm{Tr}(T \Theta T)]^{\frac{1}{2}} \ . & (d) \end{cases} \tag{2.19}$$

Preuve. (cfr. **Annexe III.7**) □.

Proposition 2.3.1 *Soient $X \in \mathcal{S}_n$, $\varepsilon \geq 0$ et $\eta \geq 0$. Alors, $\forall\, Z \in \partial_\eta \lambda_1(X), \exists\, G_\varepsilon \in \delta_\varepsilon \lambda_1(X)$ et cinq matrices $n \times n$ $(E_\varepsilon, F_\varepsilon, \Sigma, T, \Theta)$ telles que :*

$$\begin{cases} (E_\varepsilon, F_\varepsilon, \Sigma, T) \text{ satisfont } (2.19a), (2.19b), (2.19c) \text{ et } (2.19d) & (a) \\ \Theta = \mathrm{diag}(\theta_1, \ldots, \theta_n) \in \mathcal{C}_n & (b) \\ Z = G_\varepsilon + (E_\varepsilon \Sigma \Theta T F_\varepsilon^T + F_\varepsilon \Sigma \Theta T E_\varepsilon^T) + (F_\varepsilon T \Theta T F_\varepsilon^T - E_\varepsilon T \Theta T E_\varepsilon^T) & (c) \\ \mathrm{Tr}(T \Theta T) \leq \frac{\eta}{\lambda_{r_\varepsilon+1}(X) - \lambda_{r_\varepsilon}(X)} = \frac{\eta}{\Delta_\varepsilon(X)} \ . & (d) \end{cases} \tag{2.20}$$

Preuve. (cfr. **Annexe III.8**) □.

2.3.3 Application du développement vertical à λ_1

Le résultat suivant affirme que $\delta_\varepsilon \lambda_1(X)$ est une bonne approximation de $\partial_\eta \lambda_1(X)$ pour η suffisamment petit, dépendant de la séparation spectrale $\Delta_\varepsilon(X)$ (2.14).

Théorème 2.3.1 *$\forall\, \varepsilon \geq 0, \eta \geq 0$ et $D \in \mathcal{S}_n$, on a*

$$\lambda'_{1,\eta}(X; D) \leq \sigma_{\delta_\varepsilon \lambda_1(X)}(D) + \rho(\eta, \varepsilon)\ \|D\| \tag{2.21}$$

ou de manière équivalente,

$$\partial_\eta \lambda_1(X) \subset \delta_\varepsilon \lambda_1(X) + B(0, \rho(\eta, \varepsilon)) \tag{2.22}$$

où

$$\rho(\eta, \varepsilon) := \left[\frac{2\eta}{\Delta_\varepsilon(X)}\right]^{1/2} + \left[\frac{2\eta}{\Delta_\varepsilon(X)}\right] \ .$$

Preuve. Soient $\varepsilon \geq 0, \eta \geq 0, D \in \mathcal{S}_n$ et considérons $Z \in \partial_\eta \lambda_1(X)$. Utilisant (2.20c) et la proposition 1.2.2 (6), nous obtenons

$$\langle Z, D\rangle = \langle G_\varepsilon, D\rangle + \langle \Sigma\Theta T, E_\varepsilon^T DF_\varepsilon + F_\varepsilon^T DE_\varepsilon\rangle + \langle T\Theta T, F_\varepsilon^T DF_\varepsilon - E_\varepsilon^T DE_\varepsilon\rangle \ .$$

Considérons le membre de droite de l'égalité, nous avons $\langle G_\varepsilon, D\rangle \leq \sigma_{\delta_\varepsilon\lambda_1(X)}(D)$ en utilisant la définition de la fonction support et le fait que $G_\varepsilon \in \delta_\varepsilon\lambda_1(X)$.

$$\langle \Sigma\Theta T, E_\varepsilon^T DF_\varepsilon + F_\varepsilon^T DE_\varepsilon\rangle \ = \ \textstyle\sum_{i=1}^n \ \sigma_i \ \theta_i t_i[\langle D, e_i f_i^T + f_i e_i^T \ \rangle] \ \text{ où}$$

$$\Sigma\Theta T = \mathrm{diag}(\sigma_1\theta_1 t_1, \ \cdots \ , \ \sigma_n\theta_n t_n) \ ,$$

en appliquant le lemme 1.2.2 . Par conséquent, par Cauchy-Schwarz, nous avons

$$\begin{aligned} \langle \Sigma\Theta T, E_\varepsilon^T DF_\varepsilon + F_\varepsilon^T DE_\varepsilon\rangle \ &= \ \sum_{i=1}^n \ \sigma_i \ \theta_i t_i[\langle D, e_i f_i^T + f_i e_i^T \ \rangle] \\ &\leq \ \sum_{i=1}^n \ \sigma_i \ \theta_i t_i \ \|D\| \ \|e_i f_i^T + f_i e_i^T\| \ . \end{aligned}$$

Or $\|e_i f_i^T + f_i e_i^T\| = \sqrt{2}$. Dès lors par la définition de trace et l'inégalité (2.19d), nous obtenons

$$\begin{aligned} \langle \Sigma\Theta T, E_\varepsilon^T DF_\varepsilon + F_\varepsilon^T DE_\varepsilon\rangle \ &\leq \ \sqrt{2} \ \|D\| \ \mathrm{Tr}(\Sigma\Theta T) \\ &\leq \ \sqrt{2} \ \|D\| \ [\mathrm{Tr}(T\Theta T)]^{\frac{1}{2}} \ . \end{aligned}$$

De plus, en appliquant le lemme 1.2.2 et Cauchy-Schwarz

$$\begin{aligned} \langle T\Theta T, F_\varepsilon^T DF_\varepsilon - E_\varepsilon^T DE_\varepsilon\rangle \ &= \ \sum_{i=1}^n \theta_i t_i^2 \ [\langle D, f_i f_i^T - e_i e_i^T \ \rangle] \\ &\leq \ \mathrm{Tr}(T\Theta T) \ \|D\| \ (\|f_i f_i^T\| + \|e_i e_i^T\|) \\ &\leq \ 2 \ \mathrm{Tr}(T\Theta T) \ \|D\| \ . \end{aligned}$$

En conclusion, par l'inégalité (2.20d) :

$$\begin{aligned} \langle Z, D\rangle \ &\leq \ \sigma_{\delta_\varepsilon\lambda_1(X)}(D) + \sqrt{2} \ \|D\| \ [\mathrm{Tr}(T\Theta T)]^{\frac{1}{2}} + 2 \ \mathrm{Tr}(T\Theta T) \ \|D\| \\ &\leq \ \sigma_{\delta_\varepsilon\lambda_1(X)}(D) + \sqrt{2} \ \|D\| \ \left[\frac{\eta}{\Delta_\varepsilon(X)}\right]^{\frac{1}{2}} + 2 \ \|D\| \ \left[\frac{\eta}{\Delta_\varepsilon(X)}\right] \ . \end{aligned}$$

Comme cette inégalité est vérifiée pour tous les Z, elle est également vérifiée pour le suprémum et par conséquent, nous avons

$$\begin{aligned}\lambda'_{1,\eta}(X;D) &\leq \sigma_{\delta_\varepsilon\lambda_1(X)}(D) + \|D\| \left[\frac{2\eta}{\Delta_\varepsilon(X)}\right]^{\frac{1}{2}} + \|D\| \frac{2\eta}{\Delta_\varepsilon(X)} \\ &\leq \sigma_{\delta_\varepsilon\lambda_1(X)}(D) + \|D\| \rho(\eta,\varepsilon) .\end{aligned}$$

□.

Nous utiliserons ce résultat sous une forme simplifiée .

Corollaire 2.3.1 *Soient $X \in \mathcal{S}_n$, $\varepsilon \geq 0$ et $\eta \in [0, \frac{\Delta_\varepsilon(X)}{2}]$. Alors pour toute matrice $D \in \mathcal{S}_n$, on a*

$$\lambda'_{1,\eta}(X;D) \leq \sigma_{\delta_\varepsilon\lambda_1(X)}(D) + \left[\frac{8\,\eta}{\Delta_\varepsilon(X)}\right]^{\frac{1}{2}} \|D\| \tag{2.23}$$

Preuve. Puisque $\eta \leq \frac{\Delta_\varepsilon(X)}{2}$,

$$0 \leq \frac{2\eta}{\Delta_\varepsilon(X)} \leq \left[\frac{2\eta}{\Delta_\varepsilon(X)}\right]^{\frac{1}{2}} .$$

Par le théorème 2.3.1, nous avons successivement

$$\begin{aligned}\lambda'_{1,\eta}(X;D) &\leq \sigma_{\delta_\varepsilon\lambda_1(X)}(D) + \|D\| \rho(\eta,\varepsilon) \\ &\leq \sigma_{\delta_\varepsilon\lambda_1(X)}(D) + \|D\| \left[\frac{2\,\eta}{\Delta_\varepsilon(X)}\right]^{\frac{1}{2}} + \|D\| \frac{2\eta}{\Delta_\varepsilon(X)} \\ &\leq \sigma_{\delta_\varepsilon\lambda_1(X)}(D) + \|D\|\, 2 \left[\frac{2\eta}{\Delta_\varepsilon(X)}\right]^{\frac{1}{2}} \\ &\leq \sigma_{\delta_\varepsilon\lambda_1(X)}(D) + \|D\| \left[\frac{8\,\eta}{\Delta_\varepsilon(X)}\right]^{\frac{1}{2}} .\end{aligned}$$

□.

Finalement, nous allons montrer que nous avons une expression simple pour la fonction support de l'élargissement du sous-différentiel $\delta_\varepsilon\lambda_1(X)$: c'est la plus grande valeur propre d'une matrice symétrique $r_\varepsilon \times r_\varepsilon$.

Proposition 2.3.2 *Soit $X \in \mathcal{S}_n$. Alors $\forall\, \varepsilon \geq 0$ et $D \in \mathcal{S}_n$,*

$$\sigma_{\delta_\varepsilon\lambda_1(X)}(D) = \lambda_1(Q_\varepsilon^T D Q_\varepsilon) . \tag{2.24}$$

Preuve. Par l'égalité (2.16), et en raisonnant de manière similaire au développement de la preuve du théorème 2.1.3, nous avons

$$\begin{aligned}\sigma_{\delta_\varepsilon\lambda_1(X)}(D) = \sigma_{\partial\lambda_1(Q_\varepsilon Q_\varepsilon^T)}(D) &= \max_{Y\in C_{r_\varepsilon}} \langle D, Q_\varepsilon Y Q_\varepsilon^T\rangle \\ &= \max_{Y\in C_{r_\varepsilon}} \langle Q_\varepsilon^T D Q_\varepsilon, Y\rangle \\ &= \lambda_1(Q_\varepsilon^T D Q_\varepsilon)\ .\end{aligned}$$

□.

Dans le paragraphe suivant, nous étendons ces résultats à la fonction composée $f := \lambda_1 \circ A$.

2.4 Composition avec un opérateur affine

Nous rappelons que $\mathcal{A}$ représente la partie linéaire de l'opérateur A et $\mathcal{A}^*$, son adjoint. L'étude du premier ordre pour la fonction composée est basée sur la règle suivante ([8], ch.11, th.3.2.1) :

$$\partial_\varepsilon f(x) = \mathcal{A}^*\partial_\varepsilon\lambda_1(A(x)) \qquad \text{pour } x \in I\!R^m \text{ et } \varepsilon \geq 0. \tag{2.25}$$

De manière analogue, on peut montrer que l'élargissement de $\partial f(x)$ est l'ensemble convexe suivant :

$$\delta_\varepsilon f(x) := \mathcal{A}^*\delta_\varepsilon\lambda_1(A(x))\ . \tag{2.26}$$

Sa fonction support associée est notée

$$\tilde{f}'_\varepsilon(x;d) := \sigma_{\delta_\varepsilon f(x)}(d). \tag{2.27}$$

Ici, nous utilisons la notation $\tilde{f}'_\varepsilon(x;d)$ pour souligner l'analogie avec la dérivée directionnelle approchée de f en x :

$$f'_\varepsilon(x;d) := \sigma_{\partial_\varepsilon f(x)}(d). \tag{2.28}$$

Appliquant l'opérateur linéaire $\mathcal{A}^*$ à chaque terme de (2.17), nous obtenons

$$\partial f(x) \subset \delta_\varepsilon f(x) \subset \partial_\varepsilon f(x), \qquad [\textit{forme géométrique}] \tag{2.29}$$

ou de manière équivalente

$$f'(x;d) \leq \tilde{f}'_\varepsilon(x;d) \leq f'_\varepsilon(x;d). \qquad [\textit{forme analytique}]$$

La qualité de cette approximation est dérivée de l'inégalité du corollaire 2.3.1.

Proposition 2.4.1 *Pour tout* $\varepsilon \geq 0$, $\eta \in [0, \frac{\Delta_\varepsilon(A(x))}{2}]$ *et* $d \in I\!R^m$, *on a*

$$\begin{aligned} f'_\eta(x;d) &\leq \tilde{f}'_\varepsilon(x;d) + \left[\frac{8\eta}{\Delta_\varepsilon(A(x))}\right]^{1/2} \|\mathcal{A}d\| \\ &\leq \tilde{f}'_\varepsilon(x;d) + \left[\frac{8\eta}{\Delta_\varepsilon(A(x))}\right]^{1/2} \kappa\|d\| \ , \end{aligned} \tag{2.30}$$

où $\kappa := \sup_{\|x\|=1} \|\mathcal{A}(x)\|$ *est la plus grande valeur singulière de* $\mathcal{A}$.

Preuve. Nous avons successivement

$$\begin{aligned} f'_\eta(x;d) &= \sigma_{\partial_\eta f(x)}(d) \\ &= \sigma_{\mathcal{A}^*\partial_\eta\lambda_1(A(x))}(d) \\ &= \sup_{y\in\mathcal{A}^*\partial_\eta\lambda_1(A(x))} \langle y, d\rangle \\ &= \sup_{Z\in\partial_\eta\lambda_1(A(x))} \langle \mathcal{A}^*Z, d\rangle \\ &= \sup_{Z\in\partial_\eta\lambda_1(A(x))} \langle Z, \mathcal{A}d\rangle \\ &= \sigma_{\partial_\eta\lambda_1(A(x))}(\mathcal{A}d) \\ &= \lambda'_{1,\eta}(A(x);\mathcal{A}d) \ . \end{aligned}$$

Par le corollaire 2.3.1, nous avons

$$\lambda'_{1,\eta}(A(x);\mathcal{A}d) \leq \sigma_{\delta_\varepsilon\lambda_1(X)}(\mathcal{A}d) + \left[\frac{8\ \eta}{\Delta_\varepsilon(X)}\right]^{1/2} \|\mathcal{A}d\| \ .$$

Par les égalités ci-dessus, nous avons également

$$\sigma_{\delta_\varepsilon\lambda_1(X)}(\mathcal{A}d) = \sigma_{\delta_\varepsilon f(x)}(d) = \tilde{f}'_\varepsilon(x;d) \ .$$

En regroupant les égalités et inégalités, nous avons

$$\tilde{f}'_\eta(x;d) \leq \tilde{f}'_\varepsilon(x;d) + \left[\frac{8\eta}{\Delta_\varepsilon(A(x))}\right]^{1/2} \|\mathcal{A}d\| \ .$$

□.

Cependant, à partir de de la proposition 2.3.2 et l'égalité (2.26), nous pouvons montrer par un raisonnement analogue que $\tilde{f}'_\varepsilon(x;d)$ est aussi une valeur propre maximale :

$$\tilde{f}'_\varepsilon(x;d) = \lambda_1(Q_\varepsilon^T(\mathcal{A}d)Q_\varepsilon). \tag{2.31}$$

En particulier, nous avons pour $\varepsilon = 0$,

$$\tilde{f}'_0(x;d) = f'(x;d) = \lambda_1(Q_1^T(\mathcal{A}d)Q_1). \tag{2.32}$$

En théorie, pour rechercher une direction d'ε-descente, nous devons utiliser $\partial_\varepsilon f(x)$ le sous-différentiel approché de la fonction f et calculer sa fonction support. Ces calculs sont relativement coûteux. C'est pourquoi nous allons utiliser l'élargissement $\delta_\varepsilon f(x)$ plutôt que $\partial_\varepsilon f(x)$. L'égalité (2.31) quantifie l'effort demandé pour évaluer $\tilde{f}'_\varepsilon(x;d)$: cela nécessite le calcul de la plus grande valeur propre d'une matrice $r_\varepsilon \times r_\varepsilon$. Le théorème suivant donne les propriétés de descente d'une direction $d \in I\!R^m$ satisfaisant en $x \in I\!R^m$

$$\tilde{f}'_\varepsilon(x;d) < 0 \ .$$

Théorème 2.4.1 *Soit $x \in I\!R^m, \varepsilon \geq 0$ et $d \in I\!R^m$ tel que $\tilde{f}_\varepsilon{}'(x;d) < 0$. Alors*

(i) d est une direction d'$\eta(x;\varepsilon)$-descente où

$$\eta(x,\varepsilon) := \left[\frac{\tilde{f}_\varepsilon{}'(x;d)}{4\kappa\|d\|}\right]^2 \Delta_\varepsilon(A(x)).$$

(ii) En plus, supposons qu'il existe $\omega \in [0,1]$, $\delta > 0$ et $\mu > 0$ tel que

$$\tilde{f}_\varepsilon{}'(x;d) \leq -\omega\|d\|^2, \tag{2.33}$$

avec $\|d\| \geq \delta$ et $\Delta_\varepsilon(A(x)) \geq \mu$, alors

$$\eta(x,\varepsilon) \geq \left[\frac{\omega\delta}{4\kappa}\right]^2 \mu. \tag{2.34}$$

Preuve.

(i) Montrons que d est une direction d'$\eta(x;\varepsilon)$-descente, c'est-à-dire

$$f'_{\eta(x,\varepsilon)}(x,d) < 0.$$

Par l'inégalité (2.30), nous pouvons écrire

$$\begin{aligned} f'_{\eta(x,\varepsilon)}(x;d) &\leq \tilde{f}_\varepsilon{}'(x;d) + \left[8\left(\frac{\tilde{f}_\varepsilon{}'(x;d)}{4\,\kappa\|d\|}\right)^2\left(\frac{\Delta\varepsilon(A(x))}{\Delta\varepsilon(A(x))}\right)\right]^{1/2}\kappa\|d\| \\ &\leq \tilde{f}_\varepsilon{}'(x;d) + \frac{\sqrt{8}}{4}|\tilde{f}_\varepsilon{}'(x;d)| \\ &\leq \tilde{f}_\varepsilon{}'(x;d) - \frac{\sqrt{8}}{4}\,\tilde{f}_\varepsilon{}'(x;d) \qquad \text{car } \tilde{f}_\varepsilon{}'(x;d) \leq 0 \\ &= \tilde{f}_\varepsilon{}'(x;d)\left[1-\frac{\sqrt{8}}{4}\right] \\ &< 0\,. \end{aligned}$$

(ii) Montrons que

$$\left(\frac{\tilde{f}_\varepsilon{}'(x;d)}{4\,\kappa\|d\|}\right)^2\Delta_\varepsilon(A(x)) \geq \left[\frac{\omega\delta}{4\kappa}\right]^2\mu\,.$$

Par hypothèse, nous avons

$$\left(\frac{\tilde{f}_\varepsilon{}'(x;d)}{4\,\kappa\|d\|}\right)^2\Delta_\varepsilon(A(x)) \geq \left(\frac{\tilde{f}_\varepsilon{}'(x;d)}{4\,\kappa}\right)^2\frac{\mu}{\|d\|^2}\,.$$

Or

$$\frac{\tilde{f}_\varepsilon{}'(x;d)}{\|d\|} \leq -\omega\|d\| \leq -\delta\omega\,.$$

Nous avons donc,

$$-\left(\frac{\tilde{f}_\varepsilon{}'(x;d)}{\|d\|}\right) \geq \omega\delta$$

et

$$\left(\frac{\tilde{f}_\varepsilon{}'(x;d)}{\|d\|}\right)^2 \geq (\omega\delta)^2\,.$$

Par conséquent

$$\eta(x;\varepsilon) \geq \frac{(\omega\delta)^2\mu}{(4\kappa)^2} = (\frac{\omega\delta}{4\kappa})^2\mu\,.$$

□.

En conclusion,

• *Soit* $d = -g$ *avec* $g \in \delta_\varepsilon f(x)$. *Si* d *satisfait (2.33) pour* $\omega = 1$, *alors nous avons*

$$d = -\text{proj}_{\delta_\varepsilon f(x)} 0 \ .$$

En effet, grâce à (2.27) et la définition de fonction support, (2.33) s'écrit

$$\begin{aligned} \tilde{f}_\varepsilon{}'(x;-g) = \sigma_{\delta_\varepsilon f(x)}(-g) &= \max_{z \in \delta_\varepsilon f(x)} \langle z, -g \rangle \\ &= \max_{z \in \delta_\varepsilon f(x)} -\langle z, g \rangle \\ &= -\inf_{z \in \delta_\varepsilon f(x)} \langle z, g \rangle \\ &\leq -\|g\|^2 \ . \end{aligned}$$

Nous obtenons donc

$$\min_{z \in \delta_\varepsilon f(x)} \langle z, g \rangle \geq \|g\|^2.$$

Par conséquent,

$$\langle z, g \rangle \geq \langle g, g \rangle \qquad \forall\, z \in \delta_\varepsilon f(x)$$

i.e

$$\langle z - g, g \rangle \geq 0 \qquad \forall z \in \delta_\varepsilon f(x),$$

ce qui revient à dire que

$$g = \text{proj}_{\delta_\varepsilon f(x)} 0.$$

• *Quand* $\omega \in]0,1[$, g *est une approximation de cette projection.*

Pour obtenir de telles directions, nous utiliserons un algorithme de séparation présenté dans le chapitre suivant .

Chapitre 3

Algorithme du premier ordre

Nous allons maintenant donner un processus itératif pour calculer des directions de η-descente utilisant l'information stockée dans $\delta_\varepsilon f(x)$, η et ε étant liés par la relation du théorème 2.4.1. Le déplacement le long de cette direction sera déterminé par une recherche linéaire dichotomique contrôlée par un critère d'arrêt de η-descente. Enfin, l'algorithme obtenu, appelé "des valeurs propres approchées", sera présenté ainsi que son analyse de convergence.

3.1 Recherche de la direction

En vertu du théorème 2.4.1, la meilleure direction de descente obtenue en utilisant $\delta_\varepsilon f(x)$ est la solution du problème suivant

$$\begin{cases} \min \tilde{f}'_\varepsilon(x;d) \\ \|d\| \leq 1 \end{cases} \tag{3.1}$$

i.e, en utilisant la définition de $\tilde{f}'_\varepsilon(x;d)$,

$$\min_{\|d\|\leq 1} \max_{s\in\delta_\varepsilon f(x)} \langle s, d\rangle.$$

En vertu du théorème du minmax, on peut permuter "min" et "max" dans ce problème. On obtient ainsi

$$\max_{s\in\delta_\varepsilon f(x)} \min_{\|d\|\leq 1} \langle s, d\rangle.$$

Comme

$$\begin{aligned} \max_{s\in\delta_\varepsilon f(x)} \min_{\|d\|\leq 1} \langle s, d\rangle &= \max_{s\in\delta_\varepsilon f(x)} \langle s, -\frac{s}{\|s\|}\rangle \\ &= \max_{s\in\delta_\varepsilon f(x)} -\|s\| \\ &= -\min_{s\in\delta_\varepsilon f(x)} \|s\|, \end{aligned}$$

la meilleure direction de descente sera $d = -\text{proj}_{\delta_\varepsilon f(x)} 0$. Ce problème de projection est en fait un problème d'optimisation quadratique sur le cône des matrices s.d.p. En effet, par les égalités (2.26) et (2.16), le problème de projection sur $\delta_\varepsilon f(x)$ est équivalent à

$$\begin{cases} \min \|\mathcal{A}^*(Q_\varepsilon Y Q_\varepsilon^T)\|^2 \\ Y \succeq 0 \\ \text{Tr}(Y) = 1 \ . \end{cases} \tag{3.2}$$

C'est un problème quadratique à contraintes convexes. La résolution exacte de ce problème étant difficile, nous allons seulement calculer une approximation g de $\text{proj}_{\delta_\varepsilon f(x)} 0$. En fait, suite au théorème 2.4.1, nous imposerons que

$$\tilde{f}'_\varepsilon(x; -g) \ \leq \ -\omega \|g\|^2$$

où $\omega \in \ ]0, 1]$ est une tolérance (un paramètre) qui contrôle la proximité de g par rapport à la projection $\text{proj}_{\delta_\varepsilon f(x)} 0$.
L'algorithme que nous présentons ici pour résoudre approximativement (3.2) est essentiellement celui proposé dans [3] et dans ([8], ch.9, paragraphe 3) pour minimiser une forme quadratique sur un ensemble convexe et est appelé "méthode support–boîte noire". Cet algorithme est en fait un algorithme de séparation.

Algorithme 3.1.1 Méthode support-boîte noire

PAS 0 : $l = 1$, *et* $s = s_1 \in \delta_\varepsilon f(x)$.

PAS 1 : *Calculer* $d_l = -\text{proj}_{P_l} 0 \quad$ *où* $\ P_l = \text{co}\{s_1 \dots s_l\}$.

PAS 2 : *Calculer* $s_{l+1} \in \delta_\varepsilon f(x)$ *tq*

$$s_{l+1}^T \, d_l = \ \sigma_{\delta_\varepsilon f(x)} \, (d_l).$$

(On veut trouver un autre point qui ne donnera pas lieu à la même projection, on va donc se déplacer dans la direction de d_l*.)*

PAS 3 : *Critère d' arrêt :* Si $\ \sigma_{\delta_\varepsilon f(x)} \, (d_l) \ \leq \ -\omega \|d_l\|^2$, *alors* STOP.
Sinon $l \leftarrow l + 1$ et aller au PAS 1. □.

Pendant le processus de séparation, un faisceau de ε-sous–gradients est engendré $\{s_1, \dots, s_k\}$. En ce sens, la méthode du premier ordre présentée dans cette section est une "méthode faisceau".

Remarque 3.1.1 *Connaissant la direction d_l et la matrice Q_ε dont les colonnes forment une base orthonormale de $E_\varepsilon(A(x))$, nous pouvons calculer le vecteur s_{l+1} de la façon suivante :*

$$s_{l+1} = \mathcal{A}^* uu^T$$

avec $u = Q_\varepsilon v \in I\!R^n$ et $v \in E_1(Q_\varepsilon^T \mathcal{A} d_l Q_\varepsilon) \subseteq I\!R^{r_\varepsilon}$ tel que $\|v\| = 1$.
En effet, remarquons d'abord que $s_{l+1} \in \delta_\varepsilon f(x)$ i.e. que $uu^T \in \delta_\varepsilon \lambda_1(A(x))$. En effet, nous avons

$$u \in E_1(Q_\varepsilon Q_\varepsilon^T)$$

car $Q_\varepsilon Q_\varepsilon^T u = Q_\varepsilon Q_\varepsilon^T Q_\varepsilon v = Q_\varepsilon v = u$ et $\lambda_1(Q_\varepsilon Q_\varepsilon^T) = 1$. Par conséquent,

$$\begin{aligned} uu^T \in \mathrm{co}\{qq^T \mid \|q\| = 1,\ q \in E_1(Q_\varepsilon Q_\varepsilon^T)\} &= \partial\lambda_1(Q_\varepsilon Q_\varepsilon^T) \\ &= \delta_\varepsilon \lambda_1(A(x)). \end{aligned}$$

Montrons ensuite que

$$s_{l+1}^T d_l = \sigma_{\delta_\varepsilon f(x)}(d_l)$$

i.e., par (2.31) que

$$s_{l+1}^T d_l = \lambda_1(Q_\varepsilon^T(\mathcal{A} d_l) Q_\varepsilon).$$

En effet, comme $v \in E_1(Q_\varepsilon^T(\mathcal{A} d_l) Q_\varepsilon)$, nous avons

$$Q_\varepsilon^T(\mathcal{A} d_l) Q_\varepsilon v = \lambda_1(Q_\varepsilon^T(\mathcal{A} d_l) Q_\varepsilon) v,$$

et donc

$$v^T Q_\varepsilon^T(\mathcal{A} d_l) Q_\varepsilon v = \lambda_1(Q_\varepsilon^T(\mathcal{A} d_l) Q_\varepsilon) v^T v,$$

ou encore

$$\langle Q_\varepsilon v v^T Q_\varepsilon^T, \mathcal{A} d_l \rangle = \lambda_1(Q_\varepsilon^T(\mathcal{A} d_l) Q_\varepsilon).$$

D'autre part,

$$\begin{aligned} s_{l+1}^T d_l &= \langle \mathcal{A}^* uu^T, d_l \rangle \\ &= \langle uu^T, \mathcal{A} d_l \rangle \\ &= \langle Q_\varepsilon v v^T Q_\varepsilon^T, \mathcal{A} d_l \rangle. \end{aligned}$$

D'où la thèse. □.

Pour démontrer la proposition 3.1.1, nous avons besoin d'un lemme démontré en annexe (**Annexe III.9**).

Lemme 3.1.1 *Soit S une partie convexe compacte de $\mathbb{R}^n$ telle que $0 \notin S$ et soit $g = \text{proj}_S 0$. Si $g \in \text{ri } S$, alors*

(a) g est orthogonal à $\text{aff} S$ c-à-d. $\langle g, s - g\rangle = 0$ pour tout $s \in S$.

(b) $p = \text{proj}_{\{s_1,\dots,s_k\}} 0 \quad \Leftrightarrow \quad p - g = \text{proj}_{\{s_1,\dots,s_k\}-g} 0$
où $\{s_1, \dots, s_k\}$ est une partie finie de S. □.

Proposition 3.1.1 *La méthode support-boîte noire en x avec $\omega \in]0,1[$ converge en un nombre fini d'itérations. Supposons que la condition de ε-stricte-complémentarité soit satisfaite en $x \in \mathbb{R}^n$*

$$(SC)_\varepsilon \qquad \text{proj}_{\delta_\varepsilon f(x)} 0 \in \text{ri}\, \delta_\varepsilon f(x),$$

alors, la convergence finie est aussi obtenue si $\omega = 1$.

Preuve. Lorsque $\omega \in]0,1[$ il suit de la proposition 3.3.3 ([8], ch.9) que $d_k \longrightarrow \hat{d} = -\text{proj}_{\delta_\varepsilon f(x)} 0$. Comme $\hat{d}$ est tel que $\sigma_{\delta_\varepsilon f(x)}(\hat{d}) \leq -\|\hat{d}\|^2 < -\omega\|\hat{d}\|^2$, nous avons par continuité qu'à partir d'un certain k_0 le test d'arrêt sera vérifié. Pour démontrer la convergence dans le cas où $\omega = 1$ nous utilisons le lemme 3.1.1. L'algorithme support-boîte noire peut alors s'écrire
PAS 1 : Calculer $d_k + g = -\text{proj}_{\{s_1,\dots,s_k\}-g} 0$.
PAS 2 : Calculer $s_{k+1} \in \delta_\varepsilon f(x)$ tel que

$$\langle s_{k+1} - g, d_k + g\rangle \;=\; \sigma_{\delta_\varepsilon f(x)-g}(d_k + g)\ .$$

Cette dernière égalité est vraie car

$$\begin{aligned}
\sigma_{\delta_\varepsilon f(x)-g}(d_k + g) &= \max_{s \in \delta_\varepsilon f(x)} \langle d_k + g, s - g\rangle \\
&= \max_{s \in \delta_\varepsilon f(x)} \langle d_k + g, s\rangle - \langle d_k + g, g\rangle \\
&= \max_{s \in \delta_\varepsilon f(x)} \langle d_k, s\rangle + \langle s_{k+1}, g\rangle - \langle d_k + g, g\rangle\ .
\end{aligned}$$

En effet, par le lemme 3.1.1 (a), $\langle g, s\rangle = \|g\|^2$ pour tout $s \in \delta_\varepsilon f(x)$ et donc $\langle g, s\rangle = \langle g, s_{k+1}\rangle$ pour tout $s \in \delta_\varepsilon f(x)$. Comme dans l'algoritme support-boîte noire

$$\sigma_{\delta_\varepsilon f(x)}(d_k) = \langle s_{k+1}, d_k\rangle\ ,$$

nous obtenons l'égalité recherchée. Posant ensuite $\bar{d}_k = d_k + g$, $\bar{s}_{k+1} = s_{k+1} - g$ et $\bar{S} = \delta_\varepsilon f(x) - g$, nous obtenons l'algorithme 3.3.1 ([8], ch.9). Comme $g \in \text{ri } \delta_\varepsilon f(x)$, nous avons $0 \in \text{ri } \bar{S}$ et donc, en vertu de la proposition 3.3.4 ([8], ch.9), $\bar{d}_k = 0$ pour un indice k fini. Nous en déduisons que $d_k = -g = -\text{proj}_{\delta_\varepsilon f(x)} 0$ pour un indice k fini. □.

3.2 Recherche linéaire

Soient $x \in I\!R^m$ et $d \in I\!R^m$ produit par la "méthode support-boîte noire", tels que $\tilde{f}'_\varepsilon(x,d) < 0$. En vertu du théorème 2.4.1 et de la définition de direction d'η-descente, la fonction objectif peut être diminuée du nombre positif $\eta(x,\varepsilon)$. Le problème à résoudre est alors de trouver un $t > 0$ tel que

$$f(x+td) \leq f(x) - \eta(x,\varepsilon). \tag{3.3}$$

Nous exposons une recherche linéaire basée sur un schéma dichotomique contrôlé par un critère d'arrêt d'η-descente. Pour une implémentation avancée, nous nous référons à ([8], point 2.1.2 du ch. 13).

Recherche linéaire.

PAS 0 : $t_L = 0$, $t_R = +\infty$ et $t_0 = 1$.

PAS 1 : Calculer $q(t) := f(x+td)$ et $q'_+(t) := f'(x+td,d)$ en utilisant (2.32).

PAS 2 : TEST DE $\eta-$DESCENTE :
Si $f(x+td) \leq f(x) - \eta(x,\varepsilon)$ est vrai, alors STOP.

PAS 3 : RECHERCHE DICHOTOMIQUE
Si $q'_+(t) > 0$, poser $t_R = t$; sinon poser $t_L = t$.
Calculer $t = \frac{t_L+t_R}{2}$ si $t_R \neq +\infty$, sinon $t > t_L$ (par exemple $t = 10t_L$). □.

Lemme 3.2.1 *Supposons que $d \in I\!R^m$ soit une direction d' ε-descente c'est-à-dire $f'_\varepsilon(x;d) < 0$. Alors l'ensemble de tous les $t > 0$ tel que $f(x+td) < f(x) - \varepsilon$ forme un intervalle non vide.*

Preuve. Par définition $f'_\varepsilon(x;d) = \inf_{t>0} \frac{f(x+td)-f(x)+\varepsilon}{t} < 0$. Par conséquent, pour t suffisamment petit,

$$f(x+td) - f(x) + \varepsilon < 0$$

i.e

$$f(x+td) < f(x) - \varepsilon.$$

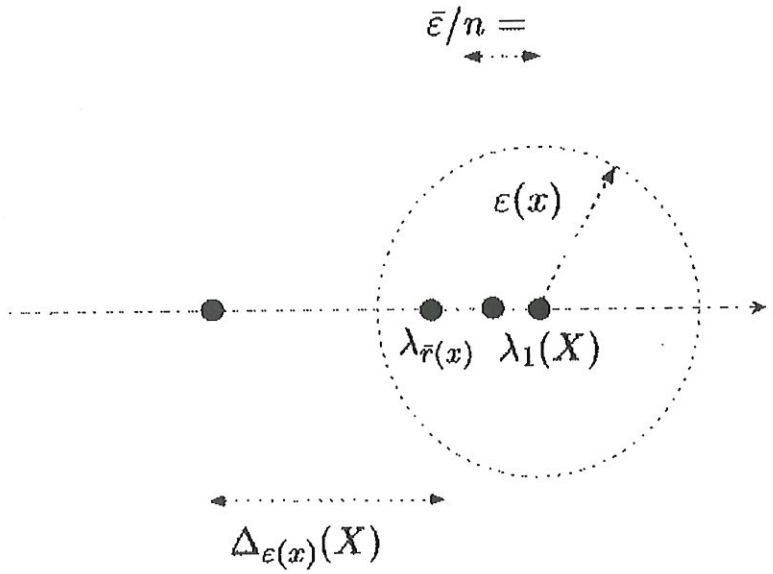
Ce qui entraîne qu'il existe bien un $t > 0$ tel que $f(x+td) < f(x) - \varepsilon$. Nous pouvons conclure au moyen de ([8], ch.13, lemme 1.2.3). □.

Théorème 3.2.1 *Soit $x \in I\!R^m$, $\varepsilon > 0$ et $d \in I\!R^m$ satisfaisant les assertions du théorème 2.4.1(ii). Alors, la recherche linéaire pour une $\eta(x,\varepsilon)-$descente s'arrête après un nombre fini d'itérations.*

Preuve. Remarquons que $\eta(x,\varepsilon) = \left[\frac{\tilde{f}_\varepsilon{}'(x;d)}{4\kappa\|d\|}\right]^2 \Delta_\varepsilon(A(x)) > 0$ car $\Delta_\varepsilon(A(x)) \geq \mu > 0$. Par le théorème 2.4.1(i), nous savons que d est une direction d'$\eta(x,\varepsilon)$-descente et donc par le lemme 3.2.1, l'ensemble de tous les $t > 0$ tel que $f(x+td) < f(x)-\varepsilon$ forme un intervalle non vide. Dès lors, la recherche dichotomique détectera un de ses éléments après un nombre fini d'opérations. □.

3.3 L'algorithme des valeurs propres approchées

L'algorithme du premier ordre que nous allons présenter ici, utilisera la ε-stratégie suivante.



Définition 3.3.1

Choisissons une tolérance $\bar{\varepsilon} > 0$. Pour $x \in \mathbb{R}^n$, définissons

$$\begin{cases} R_{\bar{\varepsilon}}(x) = \{r \in \{1, \ldots, n-1\} \ : \ \lambda_r(A(x)) - \lambda_{r+1}(A(x)) \geq \frac{\bar{\varepsilon}}{n} \\ \bar{r}(x) = \begin{cases} \{r : r \in \min R_{\bar{\varepsilon}}(x)\} & \text{si } R_{\bar{\varepsilon}}(x) \neq \emptyset \\ n & \text{sinon} \end{cases} \\ \varepsilon(x) = \begin{cases} f(x) - \lambda_{\bar{r}(x)}(A(x)) + \frac{\bar{\varepsilon}}{2n} & \text{si } R_{\bar{\varepsilon}}(x) \neq \emptyset \\ \bar{\varepsilon} & \text{sinon.} \end{cases} \end{cases} \tag{3.4}$$

Remarque 3.3.1 *En référence à la définition 2.2.1, observons que $\bar{r}(x)$ est la $\varepsilon(x)$-multiplicité de $\lambda_1(A(x))$. En effet, montrons que $\bar{r}(x) = \max\{i \mid i \in I_{\varepsilon(x)}(X)\}$. Supposons par l'absurde que $\bar{r}(x)+1$ soit le maximum de l'ensemble considéré. Si $\bar{r}(x)+1 \in I_{\varepsilon(x)}(X)$, par définition de cet ensemble d'indices (2.10), nous aurions*

$$\lambda_{\bar{r}(x)+1}(X) > \lambda_1(X) - \varepsilon(x) \ .$$

En utilisant la définition de $\varepsilon(x)$, nous obtiendrions de manière équivalente

$$\lambda_{\bar{r}(x)+1} > \lambda_{\bar{r}(x)}(X) - \frac{\bar{\varepsilon}}{2n} \ .$$

Et donc nous aurions $\lambda_{\bar{r}(x)}(X) - \lambda_{\bar{r}(x)}(X) > \frac{\bar{\varepsilon}}{n}$ ce qui contredit le fait que $\bar{r}(x)$ soit dans $R_{\bar{\varepsilon}(x)}$.
Si $R_{\bar{\varepsilon}}(x) \neq \emptyset$, la séparation spectrale à $\varepsilon(x)$-près est plus grande que $\frac{\bar{\varepsilon}}{n}$. De plus, nous avons $\lambda_r(A(x)) - \lambda_{r+1}(A(x)) < \frac{\bar{\varepsilon}}{n}$ pour $r = 1, \ldots \bar{r}(x) - 1$ quand $\bar{r}(x) > 1$.

L'algorithme suivant a pour but de minimiser f avec une tolérance $\bar{\varepsilon}$.

Algorithme 3.3.1

PAS 0 : INITIALISATION. *Choisir les tolérances $\delta > 0$ $\bar{\varepsilon} > 0$, $\omega \in]0, 1]$. Initialiser $x := x_0 \in \mathbb{R}^m$.*

PAS 1 : SÉPARATION. *Poser $\varepsilon = \varepsilon(x)$ et calculer $d \in -\delta_\varepsilon f(x)$ satisfaisant (2.33) avec la "méthode support-boîte noire".*

PAS 2 : CRITÈRE D'ARRÊT. *Si $\|d\| \leq \delta$, alors* STOP.

PAS 3 : RECHERCHE LINÉAIRE. *Calculer t tel que*

$$f(x + td) \leq f(x) - \eta(x, \varepsilon)$$

en utilisant la recherche linéaire décrite à la section précédente (backtracking, on part de $t = 1$ et on diminue t).

PAS 4 : MISE À JOUR. $x \to x + td$ *et retour en* PAS 1. □.

Pour que le problème de départ ait du sens, nous allons supposer que f soit bornée inférieurement. Nous avons

Lemme 3.3.1 *La fonction f est bornée inférieurement si et seulement si*

$$(1)\ 0 \in \mathcal{A}^*(\mathcal{C}_n)$$

ou, de manière équivalente

$$(2)\ (\mathrm{Im}A)^{\perp} \cap \mathcal{S}_n^+ \neq \{0\}.$$

Preuve. De la dualité convexe, nous avons

$$f^*(s) = \sup_x\{\langle s, x\rangle - f(x)\} = -\inf_x\{f(x) - \langle s, x\rangle\}\ .$$

Dès lors

$$-f^*(0) = \inf_{x\in I\!R^m} f(x)\ .$$

D'autre part comme (voir par exemple [22]),

$$-f^*(0) = \sup\{\langle Z, A_0\rangle :\ Z \in \mathcal{C}_n \cap\ \ker \mathcal{A}^*\},$$

nous obtenons

$$\inf_{x\in I\!R^m} f(x) = \sup\{\langle Z, A_0\rangle :\ Z \in \mathcal{C}_n \cap\ \ker \mathcal{A}^*\}.$$

Par conséquent f est bornée inférieurement si et seulement si

$$\mathcal{C}_n \cap\ \ker \mathcal{A}^* \neq \emptyset,$$

i.e.,

$$0 \in \mathcal{A}^*(\mathcal{C}_n).$$

Il reste à montrer l'équivalence entre (1) et (2).

1. Montrons la condition nécessaire : nous avons

$$0 \in \mathcal{A}^*(\mathcal{C}_n) \Leftrightarrow \exists\, Y \in \mathcal{C}_n \text{ tel que } \mathcal{A}^*(Y) = 0.$$

Par conséquent,

$$Y \in \ker \mathcal{A}^* \cap S_n^+ \ .$$

De plus $Y \neq 0$ car $Y \in \mathcal{C}_n$. Dès lors,

$$\ker \mathcal{A}^* \cap S_n^+ \neq \{0\}.$$

2. Montrons la condition suffisante. Nous savons que $\ker \mathcal{A}^* \cap S_n^+ \neq \{0\}$ i.e.

$$\exists\, \overline{Y} \neq 0 \text{ tel que } \mathcal{A}^*(\overline{Y}) = 0 \text{ et } \overline{Y} \in S_n^+.$$

Il suffit alors de prendre $Y := (\frac{1}{\sum \lambda_i})\overline{Y}$ où les λ_i sont les valeurs propres de $\overline{Y}$ pour $i = 1, \ldots, n$. □.

Nous obtenons ensuite le résultat suivant.

Lemme 3.3.2 *Pour tout vecteur $x \in \mathbb{R}^m$, le paramètre $\varepsilon(x)$ de (3.4) est inférieur ou égal à $\bar{\varepsilon}$. Dès lors,*

$$\delta_{\varepsilon(x)} f(x) \subset \partial_{\bar{\varepsilon}} f(x). \tag{3.5}$$

De plus, si f est bornée inférieurement, alors

$$\begin{array}{ll} (i) & \Delta_{\varepsilon(x)}(A(x)) \geq \dfrac{\bar{\varepsilon}}{n} \\ ou & \\ (ii) & 0 \in \delta_{\varepsilon(x)} f(x)\ . \end{array} \tag{3.6}$$

Preuve. (cfr **Annexe III.10**) □.

Nous pouvons à présent prouver la convergence finie de l'algorithme 3.3.1 vers une solution approchée de (1).

Théorème 3.3.1 *Supposons que f soit bornée inférieurement. Alors l'algorithme 3.3.1 s'arrête après un nombre fini d'itérations, fournissant $\bar{x}$ qui satisfait la condition d'optimalité approchée :*

$$f(y) \geq f(\bar{x}) - \bar{\varepsilon} - \delta \ \|y - \bar{x}\| \quad \forall y \in \mathbb{R}^m.$$

Preuve. Tant que $\|d\| > \delta$, nous avons $0 \notin \delta_{\varepsilon(x)} f(x)$ et donc en vertu du lemme 3.3.2,

$$\Delta_{\varepsilon(x)}(A(x)) \geq \frac{\bar{\varepsilon}}{n}.$$

Utilisant (2.34), cela nous donne

$$f(x) - f(x + td) \geq \eta(x, \varepsilon) \geq \left[\frac{\omega\delta}{4\kappa}\right]^2 \frac{\bar{\varepsilon}}{n}.$$

Par conséquent l'algorithme doit s'arrêter après N itérations, où N est le premier entier satisfaisant

$$f(x_0) - N \left[\frac{\omega\delta}{4\kappa}\right]^2 \frac{\bar{\varepsilon}}{n} \leq -f^*(0).$$

En effet, on cherche le premier N pour lequel nous avons

$$f(x_{N-1}) - f(x_N) < \eta(x, \varepsilon)$$

et

$$f(x_i) - f(x_{i+1}) \geq \eta(x, \varepsilon) \qquad \forall i = 0, \ldots, N-2.$$

Pour cela, réécrivons $f(x_{N-1}) - f(x_N)$. Nous avons

$$\begin{aligned} f(x_{N-1}) - f(x_N) &= f(x_0) - (f(x_0) - f(x_1)) - \ldots (\ldots - f(x_{N-1})) - f(x_N) \\ &< f(x_0) - (N-1)\,\eta(x, \varepsilon) - f(x_N). \end{aligned}$$

Si $f(x_0) - (N-1)\,\eta(x, \varepsilon) - f(x_N) < \eta(x, \varepsilon)$, alors $f(x_{N-1}) - f(x_N) < \eta(x, \varepsilon)$.
Ce qui peut encore s'écrire
si $f(x_0) - N\eta(x, \varepsilon) - f(x_N) < 0$, alors $f(x_{N-1}) - f(x_N) < \eta(x, \varepsilon)$.
Or,

$$\eta(x, \varepsilon) > \left[\frac{\omega\delta}{4\kappa}\right]^2 \frac{\bar{\varepsilon}}{n}.$$

Donc, si $f(x_0) - N \left[\frac{\omega\delta}{4\kappa}\right]^2 \frac{\bar{\varepsilon}}{n} - f(x_N) < 0$, alors $f(x_{N-1}) - f(x_N) < \eta(x, \varepsilon)$. Enfin, comme $f^*(0) = -\inf f(x) \geq -f(x_N)$, nous avons que si N vérifie

$$f(x_0) - N \left[\frac{\omega\delta}{4\kappa}\right]^2 \frac{\bar{\varepsilon}}{n} + f^*(0) \leq 0,$$

alors

$$f(x_{N-1}) - f(x_N) < \eta(x, \varepsilon).$$

La thèse est ainsi vérifiée. □.

Deuxième partie

Le deuxième ordre

Pour construire l'algorithme du second ordre, nous utilisons la théorie du $\mathcal{U}$-Lagrangien (paragraphe 4.2) faisant appel à la géométrie différentielle (paragraphe 4.1). Par la suite, nous adoptons le même procédé que celui suivi dans la première partie. En appliquant les concepts rappelés dans le chapitre 4, nous obtenons alors le développement du second ordre pour la fonction $\lambda_1(X)$ (paragraphe 5.1). Nous en déduisons des résultats analogues pour la fonction composée en introduisant la *condition de transversalité* (paragraphe 5.2). Avant de passer à la mise en place de l'algorithme du second ordre, nous illustrons le lien entre cette nouvelle condition et la différentiabilité du $\mathcal{U}$-Lagrangien. Comme attendu, nous construisons cet algorithme (paragraphes 6.1 - 6.5) à l'aide des différentes propriétés du chapitre 5. Pour terminer, nous en analysons la convergence (théorèmes 6.5.1 et 6.5.2).

Chapitre 4

Préliminaires

4.1 Rappels de géométrie différentielle

4.1.1 Définitions

Soient S une partie non vide de $\mathbb{R}^n$ et $\bar{x}$ un point de S.

• Vecteur tangent :

Un vecteur $h \in \mathbb{R}^n$ est **tangent** à S en $\bar{x}$ si

$\exists$ une suite de vecteurs de $\mathbb{R}^n$, notée $(h^k)_{k\in\mathbb{N}}$, tq $h^k \longrightarrow h$
$\exists$ une suite de réels strictement positifs, notée $(t^k)_{k\in\mathbb{N}}$, tq $t^k \longrightarrow 0$,
avec $\bar{x} + t^k h^k \in S \quad \forall\, k$.

• Cône tangent :

Le **cône tangent** à S en $\bar{x}$ est l'ensemble des vecteurs tangents à S en $\bar{x}$; il se note $\mathrm{T}_S(\bar{x})$.

• Cône normal :

Le **cône normal** à S en $\bar{x}$ est l'ensemble

$$\mathrm{N}_S(\bar{x}) = \{v \in \mathbb{R}^n \mid \langle v, h\rangle \leq 0 \quad \forall\, h \in \mathrm{T}_S(\bar{x})\} .$$

• Opérateur de classe C^r :

Une fonction $f : S \longrightarrow \mathbb{R}$ est de **classe C^r sur un ouvert S** si toutes ses dérivées partielles d'ordre r existent et sont continues sur S.

Un **opérateur** $A : S \longrightarrow \mathbb{R}^k$ est de **classe** C^r sur l'ouvert S si chacune de ses fonctions composantes est de classe C^r sur S.
Soient $\mathcal{T}$ et $\mathcal{S}$, deux espaces euclidiens, et soient $\phi : B(\hat{S}, \delta_0) \subset \mathcal{S} \longrightarrow \mathcal{T}$, où $\hat{S} \in \mathcal{S}$ et $\delta_0 \in \mathbb{R}_o^+$, un opérateur de classe C^∞. Soit aussi $\mathcal{M}$ une sous–variété de $\mathcal{S}$.

• Valeur régulière :

Nous appelons un élément Z dans $\mathcal{T}$ une **valeur régulière** de ϕ si pour chaque élément S de $\phi^{-1}(Z)$, le différentiel de ϕ en S, noté $D\phi(S)$, est surjectif. Rappelons que l'ensemble $\phi^{-1}(Z)$ est caractérisé par

$$\phi^{-1}(Z) := \{\Omega \in B(\hat{S}, \delta_0) : \phi(\Omega) = Z\} .$$

• Equation locale d'une sous–variété :

Supposons que 0 soit une valeur régulière de ϕ et $\mathcal{M} \cap B(\hat{S}, \delta_0) = \phi^{-1}(0)$. Nous appelons $\phi(S) = 0$ une **équation locale** de $\mathcal{M}$ dans $B(\hat{S}, \delta_0)$.

• Opérateur transversal :

Considérons l'opérateur A de classe C^∞ : $\mathbb{R}^m \longrightarrow \mathcal{S}$ et $\bar{x}$ un vecteur de $\mathbb{R}^m$. L'opérateur A est dit **transversal** à la sous–variété $\mathcal{M}$ en $\bar{x}$ si l'image de $\bar{x}$ par A est dans la sous–variété $\mathcal{M}$ et l'image du différentiel de A en $\bar{x}$ est transversal au sous–espace tangent à $\mathcal{M}$ en $A(\bar{x})$, i.e., si

$$A(\bar{x}) \in \mathcal{M} ,$$

et

$$\text{Im } DA(\bar{x}) + \text{T}_{\mathcal{M}}(A(\bar{x})) = \mathcal{S}.$$

4.1.2 Propriétés

Nous présentons ici quelques résultats (voir [14]) relatifs aux concepts décrits dans le paragraphe précédent. Cela nous permettra d'introduire des notations que nous utiliserons dans la suite.

Proposition 4.1.1 THEOREME DE SUBMERSION.
Soit Z une valeur régulière de ϕ. Alors l'ensemble niveau $\phi^{-1}(Z)$ est une sous–variété de $\mathcal{S}$. Pour chaque $S \in \phi^{-1}(Z)$ l'espace tangent, noté $\text{T}_{\phi^{-1}(Z)}(S)$, est le noyau du différentiel de ϕ en S :

$$\text{T}_{\phi^{-1}(Z)}(S) = \ker D\phi(S) .$$

□.

Nous pouvons en déduire le corollaire suivant.

Corollaire 4.1.1 *Soit $\phi(S) = 0$, une équation locale de $\mathcal{M}$ dans $B(\hat{S}, \delta_0)$. Pour tout élément S dans l'intersection $\mathcal{M} \cap B(\hat{S}, \delta_0)$, l'espace tangent à $\mathcal{M}$ en S vaut*

$$\mathrm{T}_{\mathcal{M}}(S) \;=\; \ker D\phi(S)\ .$$

□.

Dans notre étude du second ordre nous utiliserons un concept particulier de la géométrie différentielle à savoir celui de *paramétrisation tangentielle.* Les deux affirmations suivantes introduisent cette idée.

Proposition 4.1.2 *Soit $\phi(S) = 0$, une équation locale de la sous-variété $\mathcal{M}$ dans $B(\hat{S}, \delta_0)$. Alors il existe un scalaire $\delta \in]0, \delta_0]$, et un opérateur unique*

$$v \;:\; \mathrm{T}_{\mathcal{M}}(\hat{S}) \cap B(0,\delta) \;\to\; \mathrm{N}_{\mathcal{M}}(\hat{S})$$

tels que pour tout couple $(u,v) \in (\mathrm{T}_{\mathcal{M}}(\hat{S}),\ \mathrm{N}_{\mathcal{M}}(\hat{S}))$,

$$(\ \|u\| \le \delta,\ \|v\| \le \delta \quad \text{et} \quad \phi(\hat{S}+u+v) \;=\; 0 \quad) \;\Rightarrow\; v = v(u)\ . \tag{4.1}$$

L'opérateur v est de classe C^∞ et en $u = 0$, nous avons

$$Dv(0) \;=\; 0\ . \tag{4.2}$$

□.

Corollaire 4.1.2 *Soit $\hat{S} \in \mathcal{M}$; alors il existe $\delta > 0$ tel que*

$$\mathrm{proj}_{\mathrm{N}_{\mathcal{M}}(\hat{S})} d \;=\; v(\mathrm{proj}_{\mathrm{T}_{\mathcal{M}}(\hat{S})} d)\ , \tag{4.3}$$

pour toute direction $d \in B(0,\delta)$ satisfaisant $\hat{S} + d \in \mathcal{M}$. □.

Ce dernier corollaire affirme, en d'autres termes, que l'opérateur

$$\pi_{\hat{S}} : \mathrm{T}_{\mathcal{M}}(\hat{S}) \cap B(0,\delta) \;\ni\; u \mapsto \hat{S} + u + v(u)$$

couvre tout un voisinage de l'élément $\hat{S}$ de la sous-variété $\mathcal{M}$. Ce qui nous permet d'appeler l'opérateur $\pi_{\hat{S}}$ *paramétrisation tangentielle* de la sous-variété $\mathcal{M}$ en $\hat{S}$.

Proposition 4.1.3 *Soit $\hat{x}$ un vecteur de $A^{-1}(\mathcal{M}) \subset \mathbb{R}^n$. Si l'opérateur est transversal à la sous-variété $\mathcal{M}$ en $\hat{x}$, alors $A^{-1}(\mathcal{M})$ est une sous-variété de classe C^∞ dans un voisinage de $\hat{x}$ i.e., il existe $\rho > 0$ tel que l'intersection $B(\hat{x},\rho) \cap A^{-1}(\mathcal{M})$ est une sous-variété de classe C^∞ sur $\mathbb{R}^n$. De plus, pour tout élément de cet ensemble, on a*

$$\mathrm{T}_{A^{-1}(\mathcal{M})}(x) \;=\; [DA(x)]^{-1}\mathrm{T}_{\mathcal{M}}(A(x))\ . \tag{4.4}$$

□.

Lorsqu'un opérateur A est transversal à $\mathcal{M}$ en un point $\hat{x}$, nous pouvons donner une équation locale de $A^{-1}(\mathcal{M})$.

Proposition 4.1.4 *Soit $\hat{x} \in A^{-1}(\mathcal{M}) \subset \mathbb{R}^n$ tel que A est transversal à la sous-variété $\mathcal{M}$ en $\hat{x}$, et prenons $\phi(S) = 0$ une équation locale de la sous-variété $\mathcal{M}$ dans un voisinage de $A(\hat{x})$. Alors il existe $\rho > 0$ tel que*

(i) l'opérateur A est transversal à $\mathcal{M}$ en $x \in B(\hat{x}, \rho)$,
(ii) l'équation $\phi(A(x)) = 0$ est une équation locale de $A^{-1}(\mathcal{M}) \cap B(\hat{x}, \rho)$.

□.

4.2 Le $\mathcal{U}$-lagrangien

Soit une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ convexe et $\bar{p}$ un vecteur de $\mathbb{R}^n$.

4.2.1 La décomposition $\mathcal{U}$–$\mathcal{V}$

Lorsque nous voulons écrire le développement de Taylor d'une fonction f en un point $\bar{p}$ de non–différentiabilité, nous sommes confrontés à la non–linéarité de l'approximation du premier ordre. L'idée est alors de considérer le sous–espace vectoriel sur lequel $f(\bar{p} + .)$ est différentiable. Or nous savons (voir [18], Th.25.2) que f est différentiable en $\bar{p}$ si et seulement si $f'(\bar{p}; .)$ est linéaire. Par conséquent, nous définissons $\mathcal{U}(\bar{p})$ comme le plus grand sous–espace vectoriel de $\mathbb{R}^n$ où la dérivée directionnelle $f'(\bar{p}; .)$ est linéaire, et nous prenons le sous–espace $\mathcal{V}(\bar{p})$, son orthogonal. Ces deux sous–espaces forment une décomposition en somme directe de l'espace tout entier,

$$\mathbb{R}^n = \mathcal{U}(\bar{p}) \oplus \mathcal{V}(\bar{p}) .$$

Définition 4.2.1

La fonction $f'(\bar{p}; .)$ étant sous–linéaire, l'espace où $f'(\bar{p}; .)$ est linéaire est défini par

$$\mathcal{U}(\bar{p}) = \{d \in \mathbb{R}^n : f'(\bar{p}; d) + f'(\bar{p}; -d) = 0\}$$

et

$$\mathcal{V}(\bar{p}) = \mathcal{U}(\bar{p})^{\perp} .$$

Ces deux sous–espaces peuvent être caractérisés de plusieurs manières :

1. $\mathcal{V}_1(\bar{p})$ est le sous–espace vectoriel parallèle à l'enveloppe affine de $\partial f(\bar{p})$. Pour un élément $\bar{g}$ arbitraire dans $\partial f(\bar{p})$ nous avons

$$\mathcal{V}_1(\bar{p}) \;=\; \text{span}\{\partial f(\bar{p}) - \bar{g}\}\;,$$

et

$$\mathcal{U}_1(\bar{p}) \;=\; \mathcal{V}_1(\bar{p})^{\perp}\;,$$

i.e.,

$$d \;\in\; \mathcal{U}_1(\bar{p}) \;\Leftrightarrow\; \langle \bar{g}+v, d\rangle \;=\; \langle \bar{g}, d\rangle \quad \forall\, v \;\in\; \mathcal{V}_1(\bar{p})\;.$$

2. $\mathcal{U}_2(\bar{p})$ et $\mathcal{V}_2(\bar{p})$ sont respectivement les cônes normal et tangent à $\partial f(\bar{p})$ en un élément g^0 arbitraire qui est dans l'intérieur relatif de $\partial f(\bar{p})$.

$$\mathcal{U}_2(\bar{p}) \;=\; \mathrm{N}_{\partial f(\bar{p})}(g^0) \qquad \text{et} \qquad \mathcal{V}_2(\bar{p}) \;=\; \mathrm{T}_{\partial f(\bar{p})}(g^0)\;.$$

La proposition suivante démontre l'équivalence entre ces différentes définitions.

Proposition 4.2.1 *Considérons les trois définitions citées ci-dessus. Dans ces conditions,*

(i) $\mathcal{U}_2(\bar{p}) \;=\; \mathrm{N}_{\partial f(\bar{p})}(g^0) \;=\; \{d \;\in\; I\!R^n \;:\; \langle g - g^0, d\rangle = 0\;,\; \forall\, g \;\in\; \partial f(\bar{p})\}\;,$
et est indépendant du choix de $g^0 \;\in\; \text{ri}\;\partial f(\bar{p})$

(ii) $\mathcal{U}(\bar{p}) \;=\; \mathcal{U}_1(\bar{p}) \;=\; \mathcal{U}_2(\bar{p})\;.$

Preuve. (cfr. **Annexe III.11.**) □.

4.2.2 Définitions

Définition 4.2.2

Soit $f : I\!R^m \longrightarrow I\!R$ *une fonction convexe. Pour un* $x \in I\!R^m$, *on a la décomposition de l'espace* $I\!R^m = \mathcal{U}(x) \oplus \mathcal{V}(x)$. *Etant donné une fonction* $g \in \partial f(x)$, *on définit le* ***$\mathcal{U}$-lagrangien*** *de* f *en* x *par*

$$\mathcal{U}(x) \ni u \longmapsto L_{\mathcal{U}}(u) := \min_{v\in\mathcal{V}(x)} \{f(x+u+v) - \langle \text{proj}_{\mathcal{V}(x)} g, v\rangle_{\mathcal{V}(x)}\}. \tag{4.5}$$

Associé à (4.5), on a l' ***ensemble des minimisants*** *(qui peut être vide),*

$$\mathcal{U}(x) \ni u \hookrightarrow v(x,g;u) := \text{Argmin}_{v\in\mathcal{V}(x)}\{f(x+u+v) - \langle \text{proj}_{\mathcal{V}(x)} g, v\rangle_{\mathcal{V}(x)}\} \subset \mathcal{V}(x). \tag{4.6}$$

4.2.3 Propriétés

Nous énonçons quelques propriétés associées à ces définitions.

Proposition 4.2.2 *La fonction $L_{\mathcal{U}}$ définie en (4.5) est convexe. De plus, si $g \in \mathrm{ri}\partial f(x)$, l'ensemble $v(x,g;u)$ décrit en (4.6) est non vide et les propriétés suivantes sont satisfaites :*

(i) Le sous-différentiel de $L_{\mathcal{U}}(u)$ en u à la forme suivante

$$\partial L_{\mathcal{U}}(u) = \mathrm{proj}_{\mathcal{U}(x)}[\partial f(x+u+v) \cap (g + \mathcal{U}(x))], \tag{4.7}$$

où v est pris arbitrairement dans $v(x,g;u)$.

(ii) Quand $u = 0$, on a $v(x,g;0) = \{0\}$ et $L_{\mathcal{U}}(0) = f(x)$. De plus, $L_{\mathcal{U}}$ est différentiable en 0 et

$$\nabla L_{\mathcal{U}}(0) = \mathrm{proj}_{\mathcal{U}(x)} g. \tag{4.8}$$

□.

Proposition 4.2.3 *Avec les notations ci-dessus, on a les propriétés suivantes*

(i) La multifonction $u \hookrightarrow \partial L_{\mathcal{U}}(u)$ est continue en $u = 0$ et

$$\lim_{u \to 0} \partial L_{\mathcal{U}}(u) = \{\nabla L_{\mathcal{U}}(0)\}. \tag{4.9}$$

(ii) Pour tout $u \in \mathcal{U}(x)$, on a

$$\partial f(x+u+v) \cap (g + \mathcal{U}(x)) = \partial L_{\mathcal{U}}(u) \oplus \{\mathrm{proj}_{\mathcal{U}(x)} g\} \qquad \forall\, v \in v(x,g;u). \tag{4.10}$$

(iii) La multifonction $u \hookrightarrow \partial f(x+u+v(x,g;u)) \cap (g + \mathcal{U}(x))$ est continue en 0 et

$$\lim_{u \to 0} \partial f(x+u+v(x,g;u)) \cap (g + \mathcal{U}(x)) = \{g\}. \tag{4.11}$$

□.

Proposition 4.2.4 *Pour tout élément $u \in \mathcal{U}(x)$, on a*

$$\sup_{v \in v(x,g;u)} \|v\| = o(\|u\|). \tag{4.12}$$

□.

Proposition 4.2.5 *La multifonction $u \hookrightarrow v(x,g;u)$ est continue en $u = 0$:*

$$\lim_{u \to 0} v(x,g;u) = \{0\}. \tag{4.13}$$

□.

Chapitre 5

Analyse du second ordre

Très souvent en optimisation, une analyse du second ordre est utile pour améliorer la convergence des algorithmes. Vu la non–différentiabilité de la fonction λ_1, nous allons exploiter sa structure particulière pour obtenir une différentiabilité locale de λ_1. Nous montrons que λ_1 admet un développement du second ordre sur l'ensemble

$$\mathcal{M}_r = \{X \in \mathcal{S}_n : \lambda_1(X) = \lambda_2(X) = \ldots = \lambda_r(X) > \lambda_{r+1}(X)\},$$

où r désigne la multiplicité de $\lambda_1(X)$. La première section de ce chapitre étudie ce comportement au second ordre via le $\mathcal{U}$-Lagrangien.
Ensuite, afin d'obtenir des résultats similaires pour la fonction $f = \lambda_1 \circ A$, nous devrons prendre certaines précautions car l'intersection de la sous-variété $\mathcal{M}_r$ avec l'image de l'opérateur affine A présente certaines singularités. Pour éviter celles-ci, il faudra imposer une condition bien particulière, la *condition de transversalité.* Ceci fait l'objet de la seconde section.

5.1 Le $\mathcal{U}$-Lagrangien de λ_1

Appliquons les résultats exposés dans le chapitre précédent. Soit $X \in \mathcal{S}_n$, pour cela notons $\mathcal{U}(X)$, le plus grand sous–espace de $\mathcal{S}_n$ où la dérivée directionnelle $\lambda_1'(X,.)$ est linéaire et $\mathcal{V}(X)$ l'orthogonal de $\mathcal{U}(X)$. Nous avons

$$\mathcal{U}(X) = \{U \in \mathcal{S}_n : \lambda_1'(X;U) + \lambda_1'(X;-U) = 0\}. \tag{5.1}$$

Appliquant la proposition 4.2.1, nous avons immédiatement que les sous-espaces $\mathcal{U}(X)$ *et* $\mathcal{V}(X)$ peuvent être caractérisés comme suit :

Proposition 5.1.1 *Soit* $X \in \mathcal{S}_n$

(i) Pour toute matrice $G \in \mathrm{ri}\partial\lambda_1(X)$, $\mathcal{U}(X)$ *et* $\mathcal{V}(X)$ *sont respectivement les cônes normal et tangent à* $\partial\lambda_1(X)$ *en* G.

(ii) $\mathcal{U}(X)$ *et* $\mathcal{V}(X)$ *sont respectivement les sous-espaces orthogonaux et parallèles à* $\mathrm{aff}\partial\lambda_1(X)$. □.

Pour atteindre le second ordre il faut tenir compte du comportement local de toutes les contraintes actives en X, c-à-d. de la multiplicité de λ_1 (stratégie adoptée aussi dans [17]). Pour cela, nous allons considérer la sous–variété définie par

$$\mathcal{M}_r = \{M \in \mathcal{S}_n : \lambda_1(M) = \ldots = \lambda_r(M) > \lambda_{r+1}(M)\}.$$

Cette sous-variété $\mathcal{M}_r$ nous permet d'introduire de nouveaux concepts et propriétés admis dans ce travail (pour les preuves nous renvoyons le lecteur à [14]).

Définition 5.1.1

Soit $M \in \mathcal{M}_r$. *On définit le* ***sous-espace*** $\boldsymbol{E_{tot}(M)}$, *comme étant le sous-espace propre de dimension* r *engendré par les vecteurs propres associés à* $\lambda_1(M)$.

Définition 5.1.2

Si $X = \sum_{i=1}^{n} \lambda_i(X) q_i q_i^T$ *est la décomposition spectrale de* X, *on définit l'****inverse de Moore-Penrose*** *de* X *par*

$$X^\dagger := \sum_{\lambda_i(X)\neq 0} \frac{1}{\lambda_i(X)} q_i q_i^T.$$

Proposition 5.1.2 *Prenons* $M \in \mathcal{M}_r$ *et choisissons une base orthonormale* $Q_1(M)$ *de* $E_1(M) = E_{tot}(M)$. *Alors il existe un* $\delta > 0$ *et une fonction* $Q_{tot} : B(M,\delta) \to \mathbb{R}^{n\times r}$ *telles que*

(i) pour toute matrice $A \in B(M,\delta)$, *les colonnes de* $Q_{tot}(A)$ *forment une base orthonormale de* $E_{tot}(A)$ *et* $Q_{tot}(M) = Q_1(M)$,

(ii) Q_{tot} *est* C^∞ *et, en particulier,*

$$DQ_{tot}(M)\cdot H = (\lambda_1(M) I_n - M)^\dagger H Q_{tot}(M) \qquad \forall\, H \in \mathcal{S}_n. \tag{5.2}$$

□.

Proposition 5.1.3 *Les fonctions* $\Lambda_{tot} : B(M,\delta) \ni A \longmapsto Q_{tot}(A)^T A Q_{tot}(A)$ *et* $\hat{\lambda} : B(M,\delta) \ni A \longmapsto \frac{1}{r}\sum_{i=1}^{r} \lambda_i(A)$ *sont* C^∞. *En particulier, pour* $A \in \mathcal{M}_r \cap B(M,\delta)$ *et pour tout* $H \in \mathcal{S}_n$,

$$D\Lambda_{tot}(A)\cdot H = Q_{tot}(A)^T H Q_{tot}(A) \tag{5.3}$$

et

$$D\hat{\lambda}(A)\cdot H = \frac{1}{r}\mathrm{Tr}(Q_{tot}(A)^T H Q_{tot}(A)). \tag{5.4}$$

□.

Proposition 5.1.4

(i) La fonction $\phi : B(M,\delta_0) \ni A \longmapsto Q_{tot}(A)^T A Q_{tot}(A) - \frac{1}{r}\mathrm{Tr}(Q_{tot}(A)^T A Q_{tot}(A)) I_r$ *appartenant à l'ensemble* $\{Z \in \mathcal{S}_r : \mathrm{Tr}Z = 0\}$ *est* C^∞ *; en particulier, pour tout* $A \in \mathcal{M}_r \cap B(M,\delta_0)$*, on a*

$$D\phi(A)\cdot H = Q_{tot}(A)^T H Q_{tot}(A) - \frac{1}{r}\mathrm{Tr}(Q_{tot}(A)^T H Q_{tot}(A)) I_r \qquad \forall\, H \in \mathcal{S}_n. \tag{5.5}$$

(ii) L'équation $\phi(A) = 0$ *est une équation locale de la sous-variété* $\mathcal{M}_r$ *sur* $B(M,\delta_0)$*, et pour tout* $A \in \mathcal{M}_r \cap B(M,\delta_0)$*, on a*

$$\mathrm{T}_{\mathcal{M}_r}(A) = \ker D\phi(A). \tag{5.6}$$

□.

Ces différents résultats techniques nous permettent, entre autre, d'établir une interprétation géométrique des sous-espaces $\mathcal{U}(X)$ et $\mathcal{V}(X)$.

Théorème 5.1.1 *Soit* $X \in \mathcal{M}_r$*. Les sous-espaces* $\mathcal{U}(X)$ *et* $\mathcal{V}(X)$ *sont respectivement les espaces tangent et normal à* $\mathcal{M}_r$ *en* X *:*

$$\mathcal{U}(X) = \{U \in \mathcal{S}_n : Q_1^T U Q_1 - \frac{1}{r}\mathrm{Tr}(Q_1^T U Q_1) I_r = 0\} \tag{5.7}$$

$$\mathcal{V}(X) = \{Q_1 Y Q_1^T : Y \in \mathcal{S}_r, \langle Y, I_r\rangle = 0\}\ . \tag{5.8}$$

Preuve.

(i) L'égalité (5.7) est vérifiée. En effet, prenons un élément de $\mathrm{ri}\partial\lambda_1(X)$ i.e. par exemple son centre $C_r = \frac{1}{r}Q_1 Q_1^T$. Par la proposition 5.1.1 , $\mathcal{U}(X)$ est le cône normal de $\partial\lambda_1(X)$ en C_r . Cela signifie que

$$\begin{aligned}
U \in \mathcal{U}(X) &\Leftrightarrow \forall\, Y \in \mathcal{Z} = \{Y \in S_r^+, \mathrm{Tr}Y = 1\}, \quad 0 \geq \langle U, Q_1 Y Q_1^T - C_r\rangle \\
&\Leftrightarrow \max_{Y\in\mathcal{Z}} \langle Q_1^T U Q_1, Y\rangle \leq \frac{1}{r}\mathrm{Tr}(Q_1^T U Q_1) \\
&\Leftrightarrow \sigma_{\mathcal{Z}}(Q_1^T U Q_1) \leq \frac{1}{r}\mathrm{Tr}(Q_1^T U Q_1) \\
&\Leftrightarrow \lambda_1(Q_1^T U Q_1) \leq \frac{1}{r}\mathrm{Tr}(Q_1^T U Q_1) \\
&\Leftrightarrow r\lambda_1(Q_1^T U Q_1) \leq \mathrm{Tr}(Q_1^T U Q_1)\ .
\end{aligned}$$

Dès lors , en notant $\lambda_i = \lambda_i(Q_1^T U Q_1)$ pour $i = 1, \ldots, r$, nous avons que

$$\lambda_1 + \ldots + \lambda_r \leq r\lambda_1 \leq \lambda_1 + \ldots + \lambda_r \ .$$

Par conséquent, les r premières valeurs propres de $Q_1^T U Q_1$ sont égales à $\lambda_1(Q_1^T U Q_1)$. Nous pouvons alors écrire

$$Q_1^T U Q_1 - \frac{1}{r}(\mathrm{Tr}(Q_1^T U Q_1))I_r = 0 \ .$$

(ii) Montrons la double inclusion pour obtenir l'égalité (5.8).
a) Montrons que $\mathcal{V}(X) \subseteq \{Q_1^T U Q_1 : Y \in \mathcal{S}_r, \mathrm{Tr}Y = 0\}$.
Par la proposition 5.1.1, $\mathcal{V}(X)$ est un sous–espace vectoriel parallèle à $\mathrm{aff}(\partial\lambda_1(X))$.
Comme $\partial\lambda_1(X) = \{Q_1^T U Q_1 : Y \in \mathcal{S}_r^+, \mathrm{Tr}Y = 1\}$ par le théorème 2.1.1, nous avons alors que

$$\mathcal{V}(X) \subseteq \{Q_1^T U Q_1 : Y \in \mathcal{S}_r, \mathrm{Tr}Y = 0\} \ .$$

b) Montrons l'inclusion inverse i.e. $\mathcal{V}(X) \supseteq \{Q_1^T U Q_1 : Y \in \mathcal{S}_r, \mathrm{Tr}Y = 0\}$.
Nous avons que $Q_1\{Y \in \mathcal{Z}_r : \mathrm{Tr}Y = 0\}Q_1^T \subset \mathcal{U}(X)^\perp = \mathcal{V}(X)$. Il suffit donc de montrer que $\langle Q_1 Y Q_1^T, U\rangle = 0$ pour tous les $U \in \mathcal{U}(X)$.
Or

$$\begin{aligned}
\langle Y, Q_1^T U Q_1\rangle &= \langle Y, \frac{1}{r}\mathrm{Tr}(Q_1^T U Q_1)I_r\rangle \\
&= \frac{1}{r}\mathrm{Tr}(Q_1^T U Q_1)\langle Y, I_r\rangle \\
&= 0 \quad \text{car} \quad \mathrm{Tr}Y = 0 \ .
\end{aligned}$$

Enfin, vérifions qu'en $X \in \mathcal{M}_r$, les sous-espaces $\mathcal{U}(X)$ et $\mathcal{V}(X)$ sont respectivement les espaces tangent et normal de $\mathcal{M}_r$ en X. Par construction (cfr proposition 5.1.2), $Q_{tot}(X) = Q_1(X)$. Par le théorème 5.1.4 (ii) ,

$$\mathrm{T}_{\mathcal{M}_r}(X) = \ker D\phi(X).$$

Par le théorème 5.1.4 (i),

$$D\phi(X)\cdot H = Q_{tot}(X)^T H Q_{tot}(X) - \frac{1}{r}\mathrm{Tr}(Q_{tot}(X)^T H Q_{tot}(X))I_r \quad \forall\ H \in \mathcal{S}_n \ .$$

Par conséquent,

$$\begin{aligned}
\mathrm{T}_{\mathcal{M}_r}(X) &= \{H \in \mathcal{S}_n : Q_1(X)^T H Q_1(X) - \frac{1}{r}\mathrm{Tr}(Q_1(X)^T H Q_1(X))I_r = 0\} \\
&= \mathcal{U}(X) \quad \text{par le théorème 5.1.1.}
\end{aligned}$$

En prenant l'orthogonalité, nous obtenons $\mathrm{N}_{\mathcal{M}_r}(X) = \mathcal{V}(X)$. □.

Comme expliqué précédemment, l'analyse du second ordre demande l'utilisation du $\mathcal{U}$-lagrangien de la fonction λ_1. Rappelons–en la définition dans le cadre de ce chapitre.

Définition 5.1.3

Soit $X \in \mathcal{S}_n$, $G \in \partial\lambda_1(X)$.
Le ***$\mathcal{U}$-Lagrangien*** *de* λ_1 *en la paire primale-duale (X,G) est la fonction*

$$\mathcal{U}(X) \ni U \longmapsto L_{\mathcal{U}}(X, G; U) = \min_{V \in \mathcal{V}(x)} \lambda_1(X + U + V) - \langle G, V \rangle.$$

On définit aussi ***l'ensemble des minimisants associé*** *par*

$$V(X, G; U) = \arg\min_{V \in \mathcal{V}(X)} \lambda_1(X + U + V) - \langle G, V \rangle.$$

Le théorème suivant découle des résultats établis dans les préliminaires. Nous l'exprimons dans notre contexte spécifique.

Théorème 5.1.2 *Soient* $(X, G) \in \mathcal{S}_n \times \partial\lambda_1(X)$. *Alors les propriétés suivantes ont lieu.*

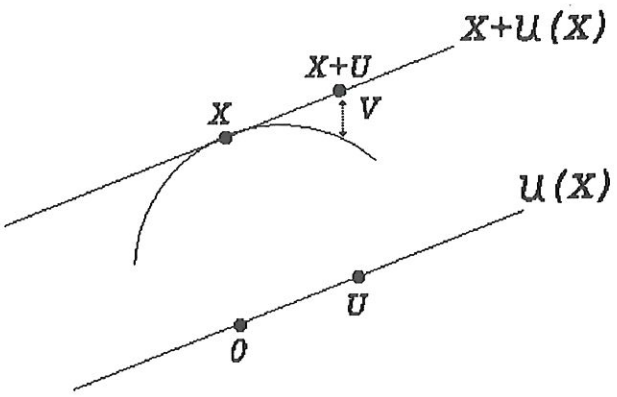
(i) La fonction $L_{\mathcal{U}}(X, G; .)$ *est bien définie et convexe sur* $\mathcal{U}(X)$
Supposons en outre que $G \in \mathrm{ri}\partial\lambda_1(X)$. *Alors*

(ii) Pour tout $U \in \mathcal{U}(X), V(X, G; U)$ *est un ensemble convexe compact non vide qui satisfait*

$$\sup_{V \in V(X,G;U)} \|V\| = o(\|U\|). \tag{5.9}$$

(iii) En particulier, en $U = 0$, *nous avons* $V(U) = \{0\}$, $L_{\mathcal{U}}(X, G; 0) = \lambda_1(X)$ *et* $\nabla L_{\mathcal{U}}(X, G; 0) = \mathrm{proj}_{\mathcal{U}(X)} G$ *existe.*

Nous pouvons interpréter géométriquement (5.9) comme suit :



Dans notre contexte, nous pouvons prouver que l'ensemble géométrique coïncide en un voisinage de X avec $\mathcal{M}_r$ quand $G \in \mathrm{ri}\partial\lambda_1(X)$:

Théorème 5.1.3 *Supposons* $(X, G) \in \mathcal{S}_n \times \mathrm{ri}\partial\lambda_1(X)$. *Alors* $\exists\ \delta > 0$ *tel que* $\forall\ U \in \mathcal{U}(X) \cap B(0, \delta)$, $V(X, G; U)$ *est un singleton.*

$$V(X, G; U) \ = \ \{V(U)\} \ .$$

De plus, l'opérateur

$$\mathcal{U}(X) \cap B(0, \delta) \ni U \longmapsto X + U + V(U) \tag{5.10}$$

est une paramétrisation C^∞ *de la surface* $\mathcal{M}_r$.

Preuve. (cfr. **Annexe III.12**). □.

Nous pouvons maintenant donner un développement du second-ordre de λ_1 le long de $\mathcal{M}_r$.

Théorème 5.1.4 *Prenons $(X,G) \in \mathcal{S}_n \times \mathrm{ri}\partial\lambda_1(X)$. Alors, $L_{\mathcal{U}}(X,G;.)$ est C^∞ dans un voisinage de $U=0$ et le développement du second ordre suivant de λ_1 a lieu :*

$$\lambda_1(X+U+V(U)) = \lambda_1(X) + \langle G, U+V(U)\rangle + \frac{1}{2}\langle \nabla^2 L_{\mathcal{U}}(X,G;0)\cdot U, U\rangle + o(\|U\|^2) \tag{5.11}$$

où $\nabla^2 L_{\mathcal{U}}(X,G;0)$ est donné explicitement par :

$$\nabla^2 L_{\mathcal{U}}(X,G;0) = \mathrm{proj}_{\mathcal{U}(X)} H(X,G)\ \mathrm{proj}^*_{\mathcal{U}(X)}$$

et $H(X,G)$ est l'opérateur symétrique s.d.p.

$$\mathcal{S}_n \ni Y \longmapsto H(X,G)Y := GY[\lambda_1(X)I_n - X]^\dagger + [\lambda_1(X)I_n - X]^\dagger YG\ .$$

L'opérateur $\nabla^2 L_{\mathcal{U}}(X,G;0)$ est appelé le $\mathcal{U}$-Hessien de λ_1 en (X,G).

Preuve. Nous utilisons les preuves des théorème 4.12 et corollaire 4.13 de [14] pour établir ce résultat. Montrons tout d'abord que le $\mathcal{U}$-Lagrangien $L_{\mathcal{U}}(X,G;.)$ est C^∞ dans un voisinage de $U=0$. Comme $\lambda_1(M) = \hat{\lambda}(M)$, où $\hat{\lambda}(M)$ est défini dans la proposition 5.1.3 pour toute matrice $M \in \mathcal{M}_r$ assez proche de X, en réunissant le théorème 5.1.3 avec la définition 5.1.3 pour $L_{\mathcal{U}}(X,G;.)$, nous écrivons le développement de $L_{\mathcal{U}}(X,G;.)$ autour de $U=0$.

$$L_{\mathcal{U}}(X,G,U) = \hat{\lambda}(\pi_X(U)) - \langle \mathrm{proj}_{\mathcal{V}(X)} G, V(U)\rangle_{\mathcal{V}(X)} \quad \forall\, U \in B(0,\rho) \tag{5.12}$$

où $\rho := \min\{\eta_{[th.5.1.3]}, \delta_{[cor.4.1.2]}, \delta_{[cor.5.1.3]}\}$.

Alors $L_{\mathcal{U}}(X,G;.)$ est C^∞ sur $B(0,\rho)$, et son développement du second ordre peut s'écrire

$$\begin{aligned} L_{\mathcal{U}}(X,G,U) &= L_{\mathcal{U}}(X,G,0) + \langle \nabla L_{\mathcal{U}}(X,G,0), U\rangle \\ &+ \frac{1}{2}\langle \nabla^2 L_{\mathcal{U}}(X,G,0)\cdot U, U\rangle + o(\|U\|^2) \\ &= \hat{\lambda}(\pi_X(U)) - \langle \mathrm{proj}_{\mathcal{V}(X)} G, V(U)\rangle_{\mathcal{V}(X)}\ . \end{aligned}$$

Nous devons encore examiner le gradient et le hessien avant d'établir le résultat annoncé pour λ_1.
(i) Nous commençons par l'étude du gradient de $L_{\mathcal{U}}$. Les résultats (2.4)-(2.5) et

(4.7) donnent

$$\nabla L_{\mathcal{U}}(U) = \mathrm{proj}_{\mathcal{U}(X)} \quad \{Q_1(\pi_X(U))Z(U)Q_1(\pi_X(U))^T \text{ tel que } Q_1(\pi_X(U))Z(U)Q_1(\pi_X(U))^T - G \in \mathcal{U}(X) \; Z \in \mathcal{S}_r^+ \,,\, \mathrm{Tr}Z = 1 \,\}.$$

Par l'égalité (5.6) et le théorème 5.1.1, le gradient de $L_{\mathcal{U}}$ s'écrit

$$\nabla L_{\mathcal{U}}(U) = \mathrm{proj}_{\mathcal{U}(X)} \quad \{Q_1(\pi_X(U))Z(U)Q_1(\pi_X(U))^T \text{ tel que } D\phi(X)\cdot(Q_1(\pi_X(U))ZQ_1(\pi_X(U))^T - G) = 0 \; (5.13) \text{ et } \mathrm{Tr}Z = 1\}.$$

Considérons ensuite le changement de variable

$$Z = \frac{1}{r} I_r + \Omega$$

où $\Omega \in \mathcal{H} = \{Z \in \mathcal{S}_r \,:\, \mathrm{Tr}Z = 0\}$.
En introduisant

$$D\phi(\pi_X(U))^* : \mathcal{H} \ni \Omega \longmapsto Q_1(\pi_X(U))\Omega Q_1(\pi_X(U))^T \,,$$

et en combinant avec (5.13), nous obtenons l'égalité

$$D\phi(X) \circ D\phi(\pi_X(U))^* \cdot \Omega = D\phi(X)\cdot G + \frac{1}{r} D\phi(X)\cdot(Q_1(\pi_X(U))Q_1(\pi_X(U))^T) \tag{5.14}$$

Puisque $D\phi(X)$ est surjectif par la proposition 5.1.4 (ii), alors $D\phi(X) \circ D\phi(X)^*$ est inversible. En effet D étant un opérateur linéaire en dimension finie, la surjectivité de $D\phi(X)$ est équivalente à son injectivité.

Par continuité, l'opérateur $D\phi(X) \circ D\phi(\pi_X(U))^*$ est aussi inversible pour U assez petit. Inversons alors (5.14). Nous obtenons la solution $\Omega(U)$ et l'opérateur

$$\mathcal{U}(X) \cap B(0,\rho) \ni U \longmapsto Z(U) = \frac{1}{r} I_r + \Omega(U) \tag{5.15}$$

est un opérateur C^∞ (car $\Omega(U)$ est C^∞). La trace de Z vaut un et $Z(U)$ est symétrique. Nous obtenons ainsi que le gradient en U dans un voisinage de $U = 0$ est

$$\nabla L_{\mathcal{U}}(X, G; U) = \mathrm{proj}_{\mathcal{U}(X)} Q_1(\pi_X(U))Z(U)Q_1(\pi_X(U))^T \tag{5.16}$$

où $Z(U)$ est caractérisé par

$$\begin{cases} Z(U) \in \{Z \in \mathcal{S}_r\ ,\ \mathrm{Tr}Z = 1\} \\ Q_1(\pi_X(U))Z(U)Q_1(\pi_X(U))^T - G = 0\ . \end{cases} \tag{5.17}$$

(ii) Calculons maintenant le terme du second ordre. Pour cela, nous dérivons (5.16) en $U = 0$; puisque nous appliquons un opérateur linéaire fixé ($\mathrm{proj}_{\mathcal{U}(X)}$) à un produit de 3 matrices, nous obtenons la somme de 3 termes.

$$\nabla^2 L_{\mathcal{U}}.H \ = \ \mathrm{proj}_{\mathcal{U}(X)}(G_1 + G_2 + G_3)$$

où

$$\begin{aligned} G_1 &= [DQ_1(X)\cdot H]ZQ_1(X)^T \\ &= [\lambda_1(X)I_n - X]^\dagger HQ_1(X)ZQ_1(X)^T \quad \text{par (5.2)}, \\ G_2 &= Q_1(X)[DZ(0)\cdot H]Q_1(X)^T, \\ G_3 &= Q_1(X)Z[DQ_1(X)\cdot H]^T \\ &= Q_1(X)ZQ_1(X)^T H[\lambda_1(X)I_n - X]^\dagger \quad \text{par (5.2)}. \end{aligned}$$

Intéressons-nous au terme suivant :

$$\mathrm{proj}_{\mathcal{U}(X)}Q_1(X)[DZ(0)\cdot H]Q_1(X)^T\ .$$

Il est nul car $DZ(0)\cdot H = D\Omega(0)\cdot H \in \mathcal{H}$ par (5.15) et donc

$$Q_1(X)[DZ(0)\cdot H]Q_1(X)^T \in \mathcal{V}(X).$$

En ce qui concerne les deux autres termes (qui sont adjoints l'un de l'autre), en combinant avec (4.2), nous observons que

$$D\pi(X)(0)\cdot H = H + DV(0)\cdot H = H.$$

Nous avons ainsi que

$$\begin{aligned} \nabla^2 L_{\mathcal{U}}\cdot H \ &= \ \mathrm{proj}_{\mathcal{U}(X)}([\lambda_1(X)I_n - X]^\dagger HQ_1(X)ZQ_1(X)^T \\ &+ \ Q_1(X)ZQ_1(X)^T H[\lambda_1(X)I_n - X]^\dagger)\ . \end{aligned}$$

Il reste à s'assurer que

$$H = \mathrm{proj}^*_{\mathcal{U}(X)}\, H\ ,$$

c'est-à-dire pour toute matrice F de $\mathcal{S}_n$, l'égalité suivante est vraie :

$$\langle H, F\rangle = \langle \mathrm{proj}^*_{\mathcal{U}(X)}\, H, F\rangle\ .$$

Vérifions cette condition en utilisant la décomposition $\mathcal{S}_n = \mathcal{U}(X) \oplus \mathcal{V}(X)$ et la définition d'adjoint. Nous avons

$$\begin{aligned} \langle H, F\rangle &= \langle H, \mathrm{proj}_{\mathcal{U}(X)} F\rangle + \langle H, \mathrm{proj}_{\mathcal{V}(X)} F\rangle \\ &= \langle H, \mathrm{proj}_{\mathcal{U}(X)} F\rangle \quad \text{car } H \in \mathcal{U}(X) = (\mathcal{V}(X))^{\perp} . \end{aligned}$$

Donc,

$$\nabla^2 L_{\mathcal{U}} \cdot H = \mathrm{proj}_{\mathcal{U}(X)} \circ H(X,G) \circ \mathrm{proj}^*_{\mathcal{U}(X)} H .$$

Nous obtenons alors que le hessien de $L_{\mathcal{U}}$ en $U = 0$ est

$$\nabla^2 L_{\mathcal{U}}(X,G,0) = \mathrm{proj}_{\mathcal{U}(X)} \circ H(X,G) \circ \mathrm{proj}^*_{\mathcal{U}(X)} \tag{5.18}$$

où $H(X,G)$ est l'opérateur symétrique défini par

$$H(X,G) \cdot Y = GY[\lambda_1(X) I_n - X]^{\dagger} + [\lambda_1(X) I_n - X]^{\dagger} Y G \qquad \forall\, Y \in \mathcal{S}_n . \tag{5.19}$$

Terminons la preuve en écrivant $L_{\mathcal{U}}$ de deux manières différentes : en utilisant son développement et en utilisant l'égalité (5.12). Nous avons

$$L_{\mathcal{U}}(X,G;U) = \lambda_1(X) + \langle \nabla L_{\mathcal{U}}(X,G;0), U\rangle + \frac{1}{2}\langle \nabla^2 L_{\mathcal{U}}(X,G;0)\cdot U, U\rangle + o(\|U\|)$$

et

$$L_{\mathcal{U}}(X,G;U) = \lambda_1(X + U + V(U)) - \langle \mathrm{proj}_{\mathcal{V}(X)} G, V(U)\rangle_{\mathcal{V}(X)}.$$

En combinant ces deux expressions, nous pouvons écrire

$$\begin{aligned} \lambda_1(X+U+V(U)) &= \lambda_1(X) + \langle \nabla L_{\mathcal{U}}(X,G;0), U\rangle + \langle \mathrm{proj}_{\mathcal{V}(X)} G, V(U)\rangle \\ &\quad + \frac{1}{2}\langle \nabla^2 L_{\mathcal{U}}(X,G;0)\cdot U, U\rangle + o(\|U\|^2) \\ &= \lambda_1(X) + \langle (\nabla L_{\mathcal{U}}(X,G;0) + \mathrm{proj}_{\mathcal{V}(X)} G), (U + V(U))\rangle \\ &\quad + \frac{1}{2}\langle \nabla^2 L_{\mathcal{U}}(X,G;0)\cdot U, U\rangle + o(\|U\|^2) . \end{aligned} \tag{5.20}$$

car $U \in \mathcal{U}(X)$ et $V(U) \in \mathcal{V}(X)$ et donc

$$\langle \nabla L_{\mathcal{U}}(X,G;0), V(U)\rangle = 0 \text{ et } \langle \mathrm{proj}_{\mathcal{V}(X)} G, U\rangle = 0 .$$

En se rappelant que

$$\nabla L_{\mathcal{U}}(X,G;0) = \mathrm{proj}_{\mathcal{U}(X)} Q_1(X) Z(0) Q_1(X)^T$$

avec $Q_1(X)Z(0)Q_1(X)^T - G = 0$. La preuve est terminée car

$$\begin{aligned} \nabla L_{\mathcal{U}}(X,G;0) + \mathrm{proj}_{\mathcal{V}(X)} G &= \mathrm{proj}_{\mathcal{U}(X)} G + \mathrm{proj}_{\mathcal{V}(X)} G \\ &= G . \end{aligned}$$

□.

5.2 Composition avec un opérateur affine

Quand nous composons λ_1 avec l'opérateur affine A, nous obtenons des résultats et des interprétations géométriques similaires à ceux obtenus pour λ_1. De plus, le sous-espace où $f'(x;.)$ est linéaire et son complément orthogonal peuvent être écrits :

$$\mathcal{U}^f(x) = \mathcal{A}^{-1}(\mathcal{U}(A(x))),$$

et

$$\mathcal{V}^f(x) = \mathcal{A}^*\mathcal{V}(A(x)).$$

A nouveau, quand on se concentre sur le second ordre, la première difficulté rencontrée est que l'image inverse de $\mathcal{M}_r$, i.e.

$$\{x \in I\!R^m \ : \ A(x) \in \mathcal{M}_r\} := \mathcal{W}_r$$

peut être non-différentiable. Alors, pour simplifier l'analyse, nous imposons la condition de transversalité introduite au paragraphe précédent. Dans notre contexte, cette condition s'exprime de la façon suivante.

Définition 5.2.1 (TRANSVERSALITÉ)
*On dit que A est **transversal** à $\mathcal{M}_r$ en $x \in \mathcal{W}_r$ si*

$$(T) \qquad \mathcal{U}(A(x)) + \mathrm{Im}\mathcal{A} = \mathcal{S}_n.$$

La condition de transversalité garantit une relation de surjection entre $\partial f(x)$ et $\partial\lambda_1(A(x))$. Comme nous sommes en dimension finie et que $\mathcal{A}$ est linéaire, la surjection est équivalente à l'injection. Nous avons aussi une relation bijective.

Lemme 5.2.1 *Supposons que la transversalité ait lieu en x dans $I\!R^m$ et prenons un élément $g \in \partial f(x)$. Alors, il existe un G unique dans $\partial\lambda_1(X)$ tel que $g = \mathcal{A}^*(G)$.*
De plus, si g est dans $\mathrm{ri}\partial f(x)$, alors G est aussi dans $\mathrm{ri}\partial\lambda_1(X)$.
En outre,

$$\begin{aligned} dim\, \mathcal{V}^f &= \frac{r(r+1)}{2} - 1 \ , \\ dim\, \mathcal{U}^f &= m + 1 - \frac{r(r+1)}{2} \ . \end{aligned} \tag{5.21}$$

Preuve. (cfr. **Annexe III.13**) □.

Finalement, la condition de transversalité est suffisante pour obtenir la différentiabilité du $\mathcal{U}$-Lagrangien de f et pour calculer son Hessien via un simple développement en chaîne.

Théorème 5.2.1 *Supposons que la transversalité a lieu en $x \in \mathbb{R}^m$ et que $g \in \mathrm{ri}\partial f(x)$. Alors, le $\mathcal{U}$-Lagrangien de f en x défini par,*

$$\mathcal{U}^f(x) \ni u \longmapsto L^f_{\mathcal{U}}(x,g;u) := \min_{v \in \mathcal{V}^f(x)} f(x+u+v) - g^T v$$

est C^∞ dans un voisinage de $u = 0$.
En particulier,

$$\nabla L^f_{\mathcal{U}}(x,g;0) = \mathrm{proj}_{\mathcal{U}^f(x)}\, g\ ,$$

et

$$\nabla^2 L^f_{\mathcal{U}}(x,g;0) = \mathrm{proj}_{\mathcal{U}^f(x)}\, H^f(x,G)\, \mathrm{proj}^*_{\mathcal{U}^f(x)}$$

où $H^f(x,G) := \mathcal{A}^* H(A(x),G)\mathcal{A}$ et G est l'unique (via le lemme 5.2.1) vecteur de $\mathrm{ri}\partial\lambda_1(A(x))$ tel que $g = \mathcal{A}^*(G)$.

Preuve. Il suffit d'appliquer un raisonnement similaire à celui utilisé pour la première partie de la preuve du théorème 5.1.4. □.

L'exemple suivant montre que la condition de transversalité n'est pas nécessaire pour garantir la différentiabilité de $L^f_{\mathcal{U}}(x,g;.)$ ni de $\mathcal{W}_r$.

5.3 Exemple

Considérez l'opérateur de $\mathbb{R}^2$ vers $\mathcal{S}^3$ défini par :

$$A(x_1,x_2) := \begin{bmatrix} x_1 - x_2 & 0 & 0 \\ 0 & x_2 - x_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Nous avons $f(x) = \max\{|x_1 - x_2|, 1\}$ et en $\hat{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, nous allons montrer que la condition de transversalité n'est pas vérifiée et que

$$\partial f(\hat{x}) = \left\{ \alpha \begin{pmatrix} -1 \\ 1 \end{pmatrix} : \alpha \in [0,1] \right\}.$$

Alors $\mathcal{V}^f(0,1) = I\!R \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ est unidimensionnel, alors que la condition de transversalité impliquerait, par le lemme 5.2.1, que $\dim \mathcal{V}^f(0,1) = 2$.
Enfin $\mathcal{W}_2$ est linéaire et donc différentiable dans un voisinage de $\hat{x}$.

$$\mathcal{W}_2 = \hat{x} + \mathcal{U}^f(x) = \hat{x} + I\!R \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Détaillons ces différentes affirmations :

1. LA CONDITION DE TRANSVERSALITÉ N'EST PAS SATISFAITE.

   Montrons que $\mathcal{U}(A(x)) \cap \mathrm{Im}\mathcal{A} \neq \mathcal{S}_n$.

   - D'abord nous avons

     $$\mathcal{U}(A(x)) = \{U \in \mathcal{S}_n : Q_1^T U Q_1 - \frac{1}{r}\mathrm{Tr}(Q_1^T U Q_1) I_r = 0\}$$

     où r est la multiplicité de la plus grande valeur propre et Q_1 est une matrice dont les colonnes forment une base de l'espace propre associé à la plus grande valeur propre.
     Ici, $n = 3$. Nous avons

     $$A(0,1) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

     Les valeurs propres de cette matrice sont ontenues en résolvant

     $$\det(A(0,1) - \lambda I_3) = \begin{bmatrix} -1-\lambda & 0 & 0 \\ 0 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{bmatrix} = 0.$$

     Nous avons $\lambda_1 = \lambda_2 = 1$ et $\lambda_3 = -1$. Dans notre cas, la multiplicité r vaut 2.
     Cherchons les vecteurs propres associés.

     $$\begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

D'où

$$Q_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Dès lors,

$$\begin{aligned} Q_1^T U Q_1 &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} u_{22} & u_{23} \\ u_{32} & u_{33} \end{bmatrix}. \end{aligned}$$

où $u_{ij} = u_{ji}$ car $U \in \mathcal{S}_n$.

Un autre élément dont nous avons besoin pour constituer $\mathcal{U}(A(x))$ est $\frac{1}{r}\mathrm{Tr}(Q_1^T U Q_1) I_r$ avec $r = 2$:

$$\frac{1}{2} \begin{bmatrix} u_{22} + u_{33} & 0 \\ 0 & u_{22} + u_{33} \end{bmatrix}.$$

Un élément de $\mathcal{U}(A(x))$ s'écrit donc sous la forme

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{12} & u_{22} & 0 \\ u_{13} & 0 & u_{22} \end{bmatrix}$$

avec $u_{ij} \in I\!R$. En effet, nous devons avoir $Q_1^T U Q_1 - \frac{1}{r}\mathrm{Tr}(Q_1^T U Q_1) = 0$, c'est-à-dire :

$$\begin{bmatrix} u_{22} & u_{23} \\ u_{32} & u_{33} \end{bmatrix} - \begin{bmatrix} \frac{u_{22}+u_{33}}{2} & 0 \\ 0 & \frac{u_{22}+u_{33}}{2} \end{bmatrix} = 0.$$

Ce qui nous donne bien

$$u_{23} = u_{32} = 0 \text{ et } u_{22} = u_{33}.$$

– Regardons ensuite l'image de $\mathcal{A}$, la partie linéaire de A.

$$\mathcal{A}(x) = \begin{bmatrix} x_1 - x_2 & 0 & 0 \\ 0 & x_1 - x_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

L'image de $\mathcal{A}$ est un sous-ensemble de matrices de $\mathcal{S}_3$ qui s'écrit de la manière suivante :

$$\mathrm{Im}\,\mathcal{A} = \left\{ V \in \mathcal{S}_3 \mid \exists\ (x_1, x_2)\ \text{ tq }\ V = \begin{bmatrix} x_1 - x_2 & 0 & 0 \\ 0 & x_1 - x_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\} .$$

Nous voyons qu'un élément quelconque de $\mathcal{S}_3$ n'est pas toujours obtenu à partir de la somme de deux éléments de $\mathcal{U}(A(x))$ et de Im$\mathcal{A}$. En effet on aura toujours que l'élément de la deuxième ligne et troisième colonne ainsi que son symétrique égaleront zéro.

2. ECRIVONS LE $\partial f(\hat{x})$.

La fonction $f(x)$ étant une fonction définie comme un maximum entre deux fonctions, nous pouvons donc appliquer la règle de calcul vue au chapitre 2. Soit

$$f(x_1, x_2) = \max\{f_1(x_1, x_2), f_2(x_1, x_2)\} ,$$

alors

$$\partial f(\hat{x}_1, \hat{x}_2) = \text{ co } \cup_{i \in I(\hat{x}_1, \hat{x}_2)} \partial f_i(\hat{x}_1, \hat{x}_2) ,$$

où $I(\hat{x}_1, \hat{x}_2)$ désigne l'ensemble des indices actifs (voir chapitre 2) en $(\hat{x}_1, \hat{x}_2)$. Ici, $I(0,1) = \{1, 2\}$ et

$$f_1(x_1, x_2) = |x_1 - x_2|,$$

$$f_2(x_1, x_2) = 1.$$

De plus, $\partial f_1 = \partial \max\{x_1 - x_2\,, -x_1 + x_2\}$; dès lors nous reproduisons le même raisonnement que pour la fonction f.

Nous obtenons alors $\partial f_1(\hat{x}_1, \hat{x}_2) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ et $\partial f_2(\hat{x}_1, \hat{x}_2) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Donc,

$$\begin{aligned} \partial f(\hat{x}_1, \hat{x}_2) &= \text{co}\left\{ \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \\ &= \text{co}\left\{ \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\} \\ &= \left\{ \alpha \begin{pmatrix} -1 \\ 1 \end{pmatrix} \ \alpha \in [0,1] \right\} . \end{aligned}$$

3. ECRIVONS LE SOUS-ESPACE $\mathcal{V}^f(x)$.

Nous avons

$$\mathcal{V}^f(x) = \mathcal{A}^*\mathcal{V}(A(x))$$

et

$$\mathcal{V}(A(x)) = \{Q_1 Y Q_1^T \ : \ Y \in \mathcal{S}_2 \, , \ \mathrm{Tr}Y = 0\}.$$

Soit $Y \in \mathcal{S}_2$ avec $\mathrm{Tr}Y = 0$. L'élément correspondant de $\mathcal{V}(A(x))$ s'écrit :

$$\begin{aligned} V &= Q_1 Y Q_1^T \\ &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} \\ y_{12} & y_{22} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & y_{11} & y_{12} \\ 0 & y_{12} & y_{22} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & y_{11} & y_{12} \\ 0 & y_{12} & y_{22} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & y_{11} & y_{12} \\ 0 & y_{12} & -y_{11} \end{bmatrix}. \end{aligned}$$

Soit S une matrice de $\mathcal{S}_3$. Nous avons $v = \mathcal{A}^*S \in \mathbb{R}^2$ et par définition de l'adjoint, nous avons :

$$\langle S, \mathcal{A}x \rangle = \langle \mathcal{A}^*S, x \rangle = \langle v, x \rangle.$$

Avec $v = (v_1, v_2)$ et par la définition du produit scalaire matricielle, nous obtenons les égalités suivantes :

$$\begin{aligned} v_1x_1 + v_2x_2 &= \mathrm{Tr}\left(\begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{bmatrix} \begin{bmatrix} x_1 - x_2 & 0 & 0 \\ 0 & x_2 - x_1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \\ &= s_{11}(x_1 - x_2) + s_{22}(x_2 - x_1) \\ &= x_1(s_{11} - s_{22}) + x_2(s_{22} - s_{11}). \end{aligned}$$

Par conséquent $\mathcal{A}^*S = \begin{bmatrix} s_{11} - s_{22} \\ s_{22} - s_{11} \end{bmatrix}$.

Prenons maintenant un élément $V \in \mathcal{V}(A(x))$. Nous avons vu que V est de la forme

$$V = \begin{bmatrix} 0 & 0 & 0 \\ 0 & y_{11} & y_{12} \\ 0 & y_{12} & -y_{11} \end{bmatrix}.$$

Dès lors

$$\mathcal{A}^*V = \begin{pmatrix} -y_{11} \\ y_{11} \end{pmatrix} = \alpha \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

avec $\alpha \in I\!R$.
Nous obtenons ainsi

$$\mathcal{V}^f(0,1) = I\!R \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Cet espace est de dimension un. Cependant si nous appliquions le lemme 5.2.1 avec $r = 2$, qui nécessite la condition de transversalité, la dimension de ce sous-espace vaudrait deux. Ceci prouve encore que la condition de transversalité n'est pas satisfaite.

4. ECRIVONS LE SOUS-ESPACE $\mathcal{U}^f(x)$

Nous savons que $\mathcal{U}^f(x) = \mathcal{A}^{-1}\mathcal{U}(A(x))$.
Donc,

$$\mathcal{U}^f(x) = \{(x_1, x_2) \text{ tq } \exists\ U \in \mathcal{U}(A(x)) \mid U = \mathcal{A}(x_1, x_2)\}.$$

Il faut donc trouver les vecteurs (x_1, x_2) de $I\!R^2$ tels que

$$\begin{bmatrix} x_1 - x_2 & 0 & 0 \\ 0 & x_2 - x_1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{12} & u_{22} & 0 \\ u_{13} & 0 & u_{22} \end{bmatrix}.$$

On a $u_{22} = 0$ et donc les vecteurs concernés appartiennent à l'ensemble $\left\{\alpha \begin{pmatrix} 1 \\ 1 \end{pmatrix} \ \alpha \in I\!R\right\} = \mathcal{U}^f(x)$.

5. EXAMINONS LE SOUS-ESPACE $\mathcal{W}_2$.

Par définition, $\mathcal{W}_2 = \{(x_1, x_2) \mid A(x_1, x_2) \in \mathcal{M}_2\}$. Pour décrire $\mathcal{W}_2$, nous devons passer par l'expression de $\mathcal{M}_2$.

Nous savons que

$$\mathcal{M}_2 = \{A(x_1, x_2) \mid \lambda_1(A) = \lambda_2(A)\}.$$

Or,

$$A(x_1, x_2) = \begin{bmatrix} x_1 - x_2 & 0 & 0 \\ 0 & x_2 - x_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Les valeurs propres de cette matrice sont $x_1 - x_2$, $x_2 - x_1$ et 1.
Soit $\lambda_1(A) = x_1 - x_2 = 1$ et nous devons avoir $x_2 - x_1 < 1$; ce qui est correct car $x_2 - x_1 = -(x_1 - x_2) = -1$.

Soit $\lambda_1(A) = x_2 - x_1 = 1$ et nous devons avoir $x_1 - x_2 < 1$; ce qui est correct car $x_1 - x_2 = -(x_2 - x_1) = -1$.

Soit $\lambda_1(A) = x_2 - x_1 = x_1 - x_2$ et nous devons avoir $1 < x_2 - x_1$ et $1 < x_1 - x_2$; ce qui n'est pas possible car on a $x_2 = x_1$ et $1 < 0$ est impossible dans $\mathbb{R}$.

Ce raisonnement nous permet d'écrire $\mathcal{M}_2$:

$$\mathcal{M}_2 = \{A(x_1, x_2) \mid x_1 - x_2 = 1 \text{ ou } x_2 - x_1 = 1\}.$$

Ensuite pour qu'un vecteur $(x_1, x_2)^T$ soit dans $\mathcal{W}_2$, il faut par construction de $\mathcal{M}_2$ que ses composantes satisfassent au moins l'une des deux conditions suivantes :

$$x_1 - x_2 = 1 \quad \text{ou} \quad x_2 - x_1 = 1.$$

$\mathcal{W}_2$ est donc constituée de deux droites parallèles. Si l'on choisit un point sur la deuxième, par exemple $\hat{x} = (0, 1)^T$, nous avons

$$\mathcal{W}_2 = \hat{x} + \mathcal{U}^f(x) ,$$

et donc $\mathcal{W}_2$ est différentiable dans un voisinage de $\hat{x}$.

De plus, pour tout élément $g \in \mathrm{ri}\partial f(\hat{x})$ et $u \in \mathcal{U}^f(0, 1)$, $v(x, g; u) = \{0\}$ et $L^f_{\mathcal{U}}(x, g; u) = 1$ est trivialement deux fois différentiable.

En effet,

$$\begin{aligned} v &= \begin{pmatrix} -\alpha \\ \alpha \end{pmatrix} \quad \alpha \in \mathbb{R} \\ u &= \begin{pmatrix} \beta \\ \beta \end{pmatrix} \quad \beta \in \mathbb{R} \\ g &= \begin{pmatrix} -\gamma \\ \gamma \end{pmatrix} \quad \gamma \in]0; 1[. \end{aligned}$$

$$\begin{aligned} L_{\mathcal{U}}^{f}(\hat{x}, g; u) &= \min_{v \in \mathcal{V}^{f}(x)} f(\hat{x} + u + v) - g^{T} v \\ &= \min_{\alpha \in I\!R} f(0 + \beta - \alpha; 1 + \beta + \alpha) - (2\gamma\alpha). \end{aligned}$$

Or,

$$\begin{aligned} f(0 + \beta - \alpha; 1 + \beta + \alpha) &= \max\{|\beta - \alpha - [1 + \beta + \alpha]|, 1\} \\ &= \max\{|-2\alpha - 1|, 1\} \\ &= \max\{|1 + 2\alpha|, 1\} \ . \end{aligned}$$

Envisageons les différents cas possibles :

Si $\underline{\alpha = 0}$, $f(\hat{x} + u + v) = 1$ et donc, $L_{\mathcal{U}}^{f}(\hat{x}, g; u) = 1$.

Si $\underline{\alpha > 0 \text{ ou } \alpha < -1}$, $f(\hat{x}+u+v) = 1+2\alpha$ et donc $L_{\mathcal{U}}^{f}(\hat{x}, g; u) = 1+2\alpha(1-\gamma) > 1$ car $1 - \gamma > 0$.

Si $\underline{\alpha = -1}$, $f(\hat{x} + u + v) = 1$ et $L_{\mathcal{U}}^{f}(\hat{x}, g; u) = 1 + 2\gamma > 1$ car $0 < \gamma < 1$.

Si $\underline{-1 < \alpha < 0}$, $f(\hat{x} + u + v) = 1$ et $L_{\mathcal{U}}^{f}(\hat{x}, g; u) = 1 - 2\gamma\alpha > 1$ car $0 < \gamma < 1$.

Le minimum pour $L_{\mathcal{U}}^{f}(\hat{x}, g; u)$ est donc atteint pour $\alpha = 0$ et vaut $L_{\mathcal{U}}^{f}(\hat{x}, g; u) = 1$. Par conséquent, $v(\hat{x}, g; u) = \{0\}$.

Chapitre 6

Algorithme global du second ordre

Nous allons d'abord expliquer comment stabiliser les éléments de l'espace $\mathcal{U}$ pour garantir la convergence. Pour cela, nous allons définir un élargissement de $\mathcal{U}$. Ensuite, nous présentons un algorithme global de type Newton.

6.1 Elargissement de $\mathcal{U}$

Soit $X \in \mathcal{S}_n$. Pour définir l'ensemble $\mathcal{U}(X)$, nous avons utilisé la fonction support du sous-différentiel de $\lambda_1(X)$. Afin d'obtenir un élargissement de $\mathcal{U}(X)$, nous remplacerons le sous-différentiel par son élargissement. Nous obtenons ainsi l'ε-version de la définition (5.1).

Définition 6.1.1

Soit $\varepsilon \geq 0$ et $X \in \mathcal{S}_n$. On définit $\mathcal{U}_\varepsilon(X)$ comme étant le plus grand sous-espace où $\sigma_{\delta_\varepsilon\lambda_1(X)}(.)$ est linéaire :

$$\mathcal{U}_\varepsilon(X) := \{U \in \mathcal{S}_n : \sigma_{\delta_\varepsilon\lambda_1(X)}(U) + \sigma_{\delta_\varepsilon\lambda_1(X)}(-U) = 0\}$$

et $\mathcal{V}_\varepsilon(X) := \mathcal{U}_\varepsilon(X)^\perp$.

La proposition 5.1.1 peut être également étendue dans ce nouveau contexte :

Proposition 6.1.1 *Soit $\varepsilon \geq 0$ et $X \in \mathcal{S}_n$. Les sous-espaces $\mathcal{U}_\varepsilon$ et $\mathcal{V}_\varepsilon$ sont caractérisés comme suit.*

(i) Pour un $G_\varepsilon \in \mathrm{ri}\delta_\varepsilon\lambda_1(X)$, $\mathcal{U}_\varepsilon(X)$ et $\mathcal{V}_\varepsilon(X)$ sont respectivement le cône normal et tangent de $\delta_\varepsilon\lambda_1(X)$ en G_ε.

(ii) $\mathcal{U}_\varepsilon(X)$ *et* $\mathcal{V}_\varepsilon(X)$ *sont respectivement les sous-espaces orthogonal et parallèle à* $\mathrm{aff}\,\delta_\varepsilon\lambda_1(X)$. □.

Nous remarquerons plus tard que $\mathcal{U}_\varepsilon(X)$ est simplement l'habituel $\mathcal{U}$ en une matrice convenablement choisie, notée X_ε. Et par conséquent, toutes les interprétations géométriques du paragraphe 6.1 pourront être reproduites .

6.2 Pas dual

Le pas dual consiste à calculer une approximation de la projection de 0 sur $\delta_\varepsilon f(x)$ avec la méthode support–boîte noire développée au paragraphe 3.1. Nous obtenons ainsi $g_\varepsilon \in \delta_\varepsilon f(x)$ tel que

$$\tilde{f}'_\varepsilon(x; -g_\varepsilon) \leq \omega \|g_\varepsilon\|^2. \tag{6.1}$$

En fait la méthode support–boîte noire produit également une matrice G_ε dans $\delta_\varepsilon\lambda_1(A(x))$ telle que $g_\varepsilon = \mathcal{A}^*(G_\varepsilon)$. Le sous-gradient g_ε sera utilisé pour garantir une descente significative et la matrice G_ε sera utilisée pour calculer le Hessien du $\mathcal{U}$-Lagrangien (voir algorithme global).
Il faut cependant remarquer que pour obtenir la convergence quadratique de la méthode de Newton, nous devrons calculer la projection exacte de 0 sur $\delta_\varepsilon f(x)$. Les résultats techniques suivants montrent que c'est possible .

Lemme 6.2.1 *Supposons que la condition de transversalité* (T) *(définition 5.2.1) ainsi que l'hypothèse* $(SC)_0$ *de la proposition 3.1.1 soient satisfaites en* $x^* \in \mathbb{R}^m$. *Alors pour* $0 < \varepsilon < \Delta_0(A(x^*))$ *(définition 2.2.1), la condition de* ε*-stricte complémentarité* $(SC)_\varepsilon$ *a lieu dans un voisinage entier de* x^*. □.

Proposition 6.2.1 *Supposons que* (T) *et* $(SC)_0$ *soient satisfaites en* x^* *et prenons* $0 < \varepsilon < \Delta_0(A(x^*))$. *Alors il existe un* ρ_1 *tel que pour tout* $x \in B(x^*, \rho_1)$, *la méthode support–boîte noire avec* $\omega = 1$ *donne la projection de* 0 *sur* $\delta_\varepsilon f(x)$ *en un nombre fini de pas.*

Preuve. Il suffit de combiner la proposition 3.1.1 et le lemme 6.2.1. □.

6.3 Pas vertical

Il serait intéressant de projeter le point courant x sur la variété $\mathcal{W}_r$ de sorte que la matrice associée à x se trouve dans $\mathcal{M}_r$. Cependant même dans le cas où la condition de transversalité est satisfaite, une telle projection est difficile à obtenir. Néanmoins, dans l'espace des matrices, il est facile de calculer un point de $\mathcal{M}_r$

qui satisfait les conditions d'optimalité du premier ordre associées au problème de projection.
Considérons une décomposition spectrale de $X \in \mathcal{M}_r$:

$$X = Q_\varepsilon \Lambda_\varepsilon Q_\varepsilon^T + R_\varepsilon \Sigma_\varepsilon R_\varepsilon^T,$$

où Λ_ε et Σ_ε sont respectivement des matrices diagonales $r_\varepsilon \times r_\varepsilon$ et $(n-r_\varepsilon)\times(n-r_\varepsilon)$, et R_ε est une matrice $n\times(n-r_\varepsilon)$ dont les colonnes forment une base orthonormale de $E_\varepsilon(X)^\perp$. Les composantes de Λ_ε sont plus grandes que celles de Σ_ε. Définissons ensuite pour tout $X \in \mathcal{S}_n$,

$$\hat{\lambda}_{1,\varepsilon}(X) := \frac{1}{r_\varepsilon} \sum_{i=1}^{r_\varepsilon} \lambda_i(X),$$

et

$$X_\varepsilon := \hat{\lambda}_{1,\varepsilon}(X) Q_\varepsilon Q_\varepsilon^T + R_\varepsilon \Sigma_\varepsilon R_\varepsilon^T$$

où la matrice X_ε est une matrice associée à X et se trouve dans $\mathcal{M}_{r_\varepsilon}$. En effet, par la définition de $\hat{\lambda}_{1,\varepsilon}(X)$, les r_ε premières valeurs propres de X_ε sont identiques et strictement plus grandes que les valeurs propres de la matrice Σ_ε.
Nous montrons dans le théorème suivant que X_ε est la projection de X sur $\mathcal{M}_r$.

Théorème 6.3.1 *La matrice X_ε satisfait aux conditions d'optimalité du premier ordre associées au problème de projection*

$$\min_{M \in \mathcal{M}_{r_\varepsilon}} \|M - X\|^2 .$$

Preuve. Soit M une solution d'un tel problème. Du théorème 5.1.1, la variété $\mathcal{M}_r$ est différentiable dans un voisinage de M et son espace normal en M est $\mathcal{V}(M)$. Alors les conditions (nécessaires) d'optimalité associées au problème sont

$$M \in \mathcal{M}_{r_\varepsilon} \quad \text{et} \quad X - M \in \mathcal{V}(M) .$$

Il est clair que $X_\varepsilon \in \mathcal{M}_{r_\varepsilon}$, que

$$X - X_\varepsilon = Q_\varepsilon [\Lambda_\varepsilon - \hat{\lambda}_{1,\varepsilon}(X) I_{r_\varepsilon}] Q_\varepsilon^T,$$

et que $\mathrm{Tr}(\Lambda_\varepsilon - \hat{\lambda}_{1,\varepsilon}(X) I_{r_\varepsilon}) = 0$. Cela entraîne, par le théorème 5.1.1, que $X - X_\varepsilon \in \mathcal{V}(X_\varepsilon)$ (car $E_1(X_\varepsilon) = E_\varepsilon(X)$) et cela complète la preuve . □.

Comme annoncé, nous montrons maintenant que $\mathcal{U}_\varepsilon(X) = \mathcal{U}(X_\varepsilon)$.

Proposition 6.3.1 *Soit $\varepsilon \geq 0$ et $X \in \mathcal{S}_n$. Nous avons*

$$\delta_\varepsilon \lambda_1(X) = \partial \lambda_1(X_\varepsilon),$$

$$\mathcal{U}_\varepsilon(X) = \mathcal{U}(X_\varepsilon) \quad \text{et} \quad \mathcal{V}_\varepsilon(X) = \mathcal{V}(X_\varepsilon).$$

Preuve. Par construction, $\lambda_1(X_\varepsilon) = \hat{\lambda}_{1,\varepsilon}(X)$ et $E_1(X_\varepsilon) = E_\varepsilon(X)$. Dès lors il suit de (2.16) et (2.5) que $\delta_\varepsilon \lambda_1(X) = \partial \lambda_1(X_\varepsilon)$. Le reste de la preuve est immédiat. □.

Le pas (dans l'espace des matrices) menant de $A(x)$ à $A_\varepsilon(x) := [A(x)]_\varepsilon$ est appelé pas vertical car il a lieu dans $\mathcal{V}_\varepsilon$.

6.4 Pas tangent

Nous supposons qu'en $x \in \mathbb{R}^m$, les pas dual et vertical ont été calculés : nous avons ainsi $g_\varepsilon(x) = \mathcal{A}^*(G_\varepsilon(x)) \in \delta_\varepsilon f(x)$ et $A_\varepsilon(x) \in \mathcal{M}_{r_\varepsilon}$. Utilisant cette information, nous définissons ensuite le programme quadratique suivant,

$$\begin{cases} \min\langle G_\varepsilon(x), U\rangle + \frac{1}{2}\langle H(A_\varepsilon(x), G_\varepsilon(x))U, U\rangle \\ U \in \mathcal{U}(A_\varepsilon(x)) \\ A_\varepsilon(x) + U \in A_0 + \mathrm{Im}(\mathcal{A}), \end{cases} \tag{6.2}$$

où H est défini au théorème 5.1.4.
Lorsque $G_\varepsilon \in \mathrm{ri}\delta_\varepsilon \lambda_1(A(x))$ i.e., (en utilisant la proposition 6.3.1) lorsque $G_\varepsilon \in \mathrm{ri}\partial\lambda_1(A_\varepsilon(x))$, ce programme quadratique est équivalent de minimiser l'approximation du second ordre de $L_{\mathcal{U}}(A_\varepsilon(x), G_\varepsilon; .)$ sous la contrainte que $A_\varepsilon(x) + U$ se trouve dans l'image de la fonction affine A : l'existence d'un pas correspondant dans l'espace des variables est guaranti. Ecrit dans cet espace, le programme quadratique (6.2) prend la forme suivante

$$\begin{cases} \min \bar{g}_\varepsilon(x)^T d + \frac{1}{2} d^T H^f(x, G_\varepsilon(x)) d \\ A(x) - A_\varepsilon(x) + \mathcal{A}d \in \mathcal{U}_\varepsilon(A(x)) \ , \end{cases} \tag{6.3}$$

où $\bar{g}_\varepsilon(x) = g_\varepsilon(x) + \mathcal{A}^* H[A(x) - A_\varepsilon(x)]$. Montrons l'équivalence entre les deux problèmes. La fonction objectif de départ est

$$\langle G_\varepsilon(x), U\rangle + \frac{1}{2}\langle H(A_\varepsilon(x), G_\varepsilon(x))U, U\rangle$$

où $U = A_0 - A_\varepsilon(x) + \mathcal{A}(x+d)$. En remplaçant U par son expression nous obtenons

$$
\begin{aligned}
&\langle G_\varepsilon(x), U\rangle + \tfrac{1}{2}\langle H(A_\varepsilon(x), G_\varepsilon(x))U, U\rangle \\
= \;& \langle G_\varepsilon(x), A(x) - A_\varepsilon(x) + \mathcal{A}d\rangle \\
&+\frac{1}{2}\langle H(A_\varepsilon(x), G_\varepsilon(x))[A(x) - A_\varepsilon(x) + \mathcal{A}d], [A(x) - A_\varepsilon(x) + \mathcal{A}d]\rangle \\
= \;& \langle G_\varepsilon(x), A(x) - A_\varepsilon(x)\rangle + \langle \mathcal{A}^* G_\varepsilon(x), d\rangle \\
&+\frac{1}{2}\langle H[A(x) - A_\varepsilon(x)], A(x) - A_\varepsilon(x)\rangle + \frac{1}{2}\langle H[A(x) - A_\varepsilon(x)], \mathcal{A}d\rangle \\
&+\frac{1}{2}\langle H\mathcal{A}d, A(x) - A_\varepsilon(x)\rangle + \frac{1}{2}\langle H\mathcal{A}d, \mathcal{A}d\rangle \\
= \;& \bar{g}_\varepsilon(x)^T d + \frac{1}{2} d^T H^f(x, G_\varepsilon(x))d + \langle G_\varepsilon(x) + H[A(x) - A_\varepsilon(x)], A(x) - A_\varepsilon(x)\rangle \; .
\end{aligned}
$$

En éliminant les termes indépendants de la variable d, nous obtenons bien le problème (6.3). Pour le résoudre, nous allons supposer que le domaine admissible de ce problème est non vide c-à-d qu'il existe un $d_0 \in I\!R^m$ tel que

$$A(x) - A_\varepsilon(x) + \mathcal{A}d_0 \in \mathcal{U}_\varepsilon(A(x)). \tag{6.4}$$

Posant $u := d - d_0$, le programme (6.3) peut s'écrire

$$\begin{cases} \min b_\varepsilon(x)^T u + \frac{1}{2} u^T H^f(x, G_\varepsilon(x))u \\ u \in \mathcal{U}_\varepsilon^f(x), \end{cases} \tag{6.5}$$

où $H^f(x, G_\varepsilon(x))$ est défini au théorème 5.2.1, $b_\varepsilon(x) := \bar{g}_\varepsilon(x) + H^f(x, G_\varepsilon(x))d_0$, et $\mathcal{U}_\varepsilon^f(x) := \{u \in I\!R^m : \mathcal{A}u \in \mathcal{U}_\varepsilon(A(x))\}$.
Pour montrer l'équivalence entre les deux problèmes, examinons d'abord les fonctions objectif. Pour plus de facilités, nous noterons H^f à la place de $H^f(x, G_\varepsilon(x))$. Sachant que cette matrice est symétrique, nous avons les égalités suivantes :

$$
\begin{aligned}
b_\varepsilon(x)^T u + \frac{1}{2} u^T H^f u &= \bar{g}_\varepsilon^T(x)(d - d_0) + d_0^T (H^f)^T (d - d_0) + \frac{1}{2}(d - d_0)^T H^f (d - d_0) \\
&= \bar{g}_\varepsilon^T(x) d - g_\varepsilon^T(x) d_0 + d_0^T (H^f)^T d - d_0^T (H^f)^T d_0 + \frac{1}{2} d^T H^f d \\
&- \frac{1}{2} d_0^T H^f d - \frac{1}{2} d^T H^f d_0 + \frac{1}{2} d_0^T H^f d_0 \\
&= \bar{g}_\varepsilon^T(x) d - g_\varepsilon^T(x) d_0 - d_0^T (H^f)^T d_0 + \frac{1}{2} d^T H^f d + \frac{1}{2} d_0^T H^f d_0.
\end{aligned}
$$

Comme la minimisation dans (6.3) porte sur la variable d, les termes dépendant explicitement de d_0 sont négligeables. Par conséquent, les deux fonctions à minimiser des problèmes (6.3) et (6.5) sont identiques.
Il reste à montrer l'équivalence entre les contraintes. Pour cela, soustrayant (6.4)

de la contrainte de (6.3), nous avons $\mathcal{A}(d-d_0) \in \mathcal{U}_\varepsilon(A(x))$, car ce dernier ensemble est un sous–espace vectoriel. Dès lors, $\mathcal{A}u \in \mathcal{U}_\varepsilon(A(x))$ i.e, $u \in \mathcal{U}^f_\varepsilon(x)$.

Pour garantir que u est bien défini dans le problème (6.5), nous supposerons par la suite que $H^f(x, G_\varepsilon(x))$ est définie positive.
Enfin, pour assurer la convergence globale, nous vérifierons si la direction $d = d_0 + u$ satisfait

$$\tilde{f}'_\varepsilon(x; d) \leq -\omega' \|d^2\|, \tag{6.6}$$

où ω' est un nombre donné dans $]0, \omega[$.

6.5 Algorithme global

Nous pouvons maintenant présenter l'algorithme global de $\mathcal{U}$-Newton.

Algorithme 6.5.1

PAS 0 : INITIALISATION. *Choisir les tolérances* $\delta > 0$, $\bar{\varepsilon} > 0$, $\omega \in]0, 1]$ *et* $\omega' \in]0, \omega[$ *; initialiser* $x := x_0 \in I\!R^m$ *et poser* $\varepsilon := \varepsilon(x)$.

PAS 1 : PAS DUAL. *Calculer* $g_\varepsilon(x) \in \delta_\varepsilon f(x)$ *et* $G_\varepsilon(A(x)) \in \delta_\varepsilon \lambda_1(A(x))$ *satisfaisant (6.1) en utilisant la méthode support–boîte noire.*

PAS 2 : CRITERE D'ARRET. *Si* $\|g_\varepsilon(x)\| \leq \delta$, *alors on arrête .*

PAS 3 : PAS VERTICAL. *Calculer* $A_\varepsilon(x)$.

PAS 4 : PAS HORIZONTAL. *Si (6.4) est admissible et* $H^f(x, G_\varepsilon(x))$ *est définie positive, alors poser d la solution de (6.3). Si (6.6) est satifait et* $\|d\| > \delta$, *aller au* PAS 5. *Si l'une de ces conditions n'est pas satisfaite, poser* $d = -g_\varepsilon(x)$.

PAS 5 : RECHERCHE LINEAIRE. *Calculer t tel que*

$$f(x + td) \leq f(x) - \eta(x, \varepsilon),$$

en utilisant la recherche linéaire du paragraphe 3.2 .

PAS 6 : MISE A JOUR. *Remplacer x par* $x + td$ *et* ε *par* $\varepsilon(x + td)$ *; retourner au* PAS 1. □.

Analysons les conditions de convergence de cet algorithme.

Théorème 6.5.1 *Supposons que f soit bornée inférieurement. Alors l'algorithme 6.5.1 (avec $\omega < 1$) s'arrête après un nombre fini d'itérations, fournissant $\bar{x}$ qui satisfait la condition de minimalité approchée :*

$$f(y) \geq f(\bar{x}) - \bar{\varepsilon} - \delta\|y - \bar{x}\| \quad \forall\, y \in \mathbb{R}^m. \tag{6.7}$$

Preuve. La preuve est semblable au théorème 3.3.1, puisque (6.6) est satisfait à chaque itération . □ .

Pour obtenir la convergence quadratique, nous imposons une condition supplémentaire.

Définition 6.5.1

Soit $x^* \in \mathbb{R}^m$ une solution de (1) . On dit que la condition stricte du second ordre (SSOC) a lieu en x^* si les conditions (T) de la définition 5.2.1 et $(SC)_0$ de la proposition 3.1.1 ont lieu en x^* et si le $\mathcal{U}$-Hessien de f en $(x^*, 0)$ est défini positif.

Donnons d'abord les conséquences de (SSOC) :

Lemme 6.5.1 *Supposons que x^* soit une solution de (1) et que (SSOC) ait lieu en x^*. Alors*

(i) x^* est la solution unique de (1),

(ii) pour tout $\rho > 0$, il existe un $\alpha > 0$ tel que

$$f(x) \leq f(x^*) + \alpha \quad \Rightarrow \quad x \in B(x^*, \rho),$$

(iii) pour tout $\rho > 0$, il existe un $\bar{\varepsilon}$ et un δ assez petits tels que l'algorithme 6.5.1 fournit au moins une itération dans $B(x^*, \rho)$, et toutes les itérations suivantes restent dans $B(x^*, \rho)$.

Preuve. (cfr. **Annexe III.14.**) □.

Le lemme suivant nous permettra de garantir qu'en restant assez près de la solution x^*, la multiplicité exacte r^* est identifiée.

Lemme 6.5.2 *Supposons que* $0 < \bar{\varepsilon} < \Delta_0(A(x^*))$. *Alors, il existe* $\rho_2 > 0$ *tel que pour tout* $x \in B(x^*, \rho_2)$,

$$r_{\bar{\varepsilon}}(x) = \bar{r}(x) = r^*,$$

où $\bar{r}(x)$ *est défini en (2.30) et* $r_{\bar{\varepsilon}}(x) := \dim E_{\bar{\varepsilon}}(x)$.

Preuve. (cfr. **Annexe III.15.**) □.

Le théorème suivant démontre la convergence quadratique de l'algorithme sous certaines conditions.

Théorème 6.5.2 *Soit* ε *tel que* $0 < \varepsilon < \Delta_0(A(x^*))$. *Supposons que (1) ait une solution* $x^* \in \mathbb{R}^m$ *pour laquelle la condition (SSOC) est vérifiée et que l'algorithme 6.5.1 soit appliqué avec* $\omega = 1$ *et* $\bar{\varepsilon} > 0, \delta > 0$ *et* $\|x_0 - x^*\|$ *suffisamment petits. Supposons aussi que la solution de (6.3) soit acceptée au* PAS 4 *et que le* PAS 5 *produise* $t = 1$. *Alors, il existe* $C > 0$ *tel que*

$$\|x^+ - x^*\| \leq C\|x - x^*\|^2. \tag{6.8}$$

Preuve. Considérons $\rho_1 > 0$ donné par la proposition 6.2.1 (avec $\varepsilon = \bar{\varepsilon}$) et $\rho_2 > 0$ donné par le lemme 6.5.2 et supposons que $x_0 \in B(x^*, \rho)$ où $0 < \rho < \min\{\rho_1, \rho_2\}$. Par hypothèse, (1) admet la solution x^* et (SSOC) a lieu en x^*. Donc par le lemme 6.5.1(iii), toutes les itérations suivantes sont dans la boule $B(x^*, \rho)$. Alors, par le lemme 6.5.2, $r_{\bar{\varepsilon}}(x) = \bar{r}(x) = r^*$ et par la proposition 6.2.1, la "méthode–support boîte noire" avec $\omega = 1$ converge en un nombre fini de pas. En outre, si nous avons (6.6) et si la longueur de pas $t = 1$ est acceptée par la recherche linéaire, l'algorithme 6.5.1 coïncide avec l'algorithme local du second ordre décrit en [14]. Cet algorithme possède une convergence quadratique ([14], th.6.13) : pour ρ suffisamment petit, il existe $C > 0$ tel que l'on ait (6.8) quand $x \in B(x^*, \rho)$. □.

Troisième partie

Méthode faisceau de type proximal

Chapitre 7

Une méthode primale de type faisceau

L'algorithme du premier ordre donné dans la première partie, avait pour but de décrire un processus facilement interprétable géométriquement. Nous devons cependant noter deux éléments importants :

1. Le rapport $\frac{\text{utilisation de l'information}}{\text{coût de calcul}}$ dans l'algorithme (3.3.1) est assez bas. En effet, les sous–gradients et ε-sous–gradients calculés durant le processus global (i.e. le PAS 1) ou lors de la recherche linéaire locale ne sont utilisés qu'une seule fois.
2. L'algorithme est très sensible au choix d'une ε-stratégie; en choisissant à chaque itération $\varepsilon = \varepsilon(x)$ dans (3.4), on assure une convergence globale mais ce n'est peut être pas la meilleur politique à suivre en pratique.

Dans ce chapitre, nous allons présenter une méthode primale de type faisceau. L'algorithme obtenu dans sa version du premier ordre, est proche de celui décrit dans [10].

7.1 Meilleure utilisation de l'information

Une première étape consiste à utiliser d'anciens $\bar{\varepsilon}$-sous–gradients au point courant pour un certain $\varepsilon \geq \bar{\varepsilon}$. Cette idée peut s'exprimer comme suit :

Proposition 7.1.1 *Soit $k > 1$. Supposons qu', en chaque point x_i, on ait calculé $g_i = \mathcal{A}^* G_i \in \delta_{\bar{\varepsilon}} f(x_i)$ avec $G_i \in \mathcal{C}_n$, pour $i = 1, \ldots k-1$. Soit Q_k, une matrice $n \times r_k$ dont les colonnes forment une base orthonormale de $E_{\bar{\varepsilon}}(A(x_k))$.*
Posons $\widetilde{\alpha}_k := [\alpha_1, \ldots, \alpha_{k-1}]^T \in I\!R^{k-1}$ et $1_{k-1} \in I\!R^{k-1}$, le vecteur constitué de 1 ;

pour $\varepsilon \geq 0$, considérons aussi l'ensemble

$$\begin{aligned}\mathcal{G}_{k,\varepsilon} := \{\sum_{i=1}^{k-1} \alpha_i G_i + Q_k Y Q_k^T : \quad & \widetilde{\alpha}_k \in I\!R_+^{k-1}, Y \in \mathcal{S}_{r_k}^+ \\ & 1_{k-1}^T \widetilde{\alpha}_k \ + \ \langle I_{r_k}, Y \rangle \ = 1 \\ & \langle \sum_{i=1}^{k-1} \alpha_i G_i + Q_k Y Q_k^T, A(x_k) \rangle \ \geq \ f(x_k) - \varepsilon\}.\end{aligned}$$

Alors, pour tout $\varepsilon \geq \bar{\varepsilon}$, on a

$$\delta_{\bar{\varepsilon}} f(x_k) \ \subset \ \mathcal{A}^* \mathcal{G}_{k,\varepsilon} \ \subset \ \partial_\varepsilon f(x_k). \tag{7.1}$$

Preuve.

1. Montrons d'abord la deuxième inclusion, i.e. $\mathcal{A}^* \mathcal{G}_{k,\varepsilon} \ \subset \ \partial_\varepsilon f(x_k)$. Quel que soit l'élément G de $\mathcal{G}_{k,\varepsilon}$, son image par l'opérateur linéaire $\mathcal{A}^*$ doit se trouver dans le sous–différentiel approché $\partial_\varepsilon f(x_k)$. Notons $\overline{G}$, un élément de $\mathcal{G}_{k,\varepsilon}$, il s'écrit

$$\overline{G} \ := \ \sum_{i=1}^{k-1} \alpha_i G_i + Q_k Y Q_k^T \ ,$$

avec les propriétés

$$\begin{aligned}& \widetilde{\alpha}_k \in I\!R_+^{k-1}, Y \in \mathcal{S}_{r_k}^+ \\ & 1_{k-1}^T \widetilde{\alpha}_k \ + \ \langle I_{r_k}, Y \rangle \ = 1 \\ & \langle \sum_{i=1}^{k-1} \alpha_i G_i + Q_k Y Q_k^T, A(x_k) \rangle \ \geq \ f(x_k) - \varepsilon.\end{aligned}$$

Or, par les égalités (2.8) et (2.25),

$$\begin{aligned}\partial_\varepsilon f(x_k) \ &= \ \mathcal{A}^* \partial_\varepsilon \lambda_1(A(x_k)) \\ &= \ \mathcal{A}^* \{Z \in \mathcal{C}_n \ : \ \langle Z, A(x_k) \rangle \geq \lambda_1(A(x_k)) - \varepsilon\} \\ &= \ \mathcal{A}^* \{Z \in \mathcal{C}_n \ : \ \langle Z, A(x_k) \rangle \ \geq \ f(x_k) - \varepsilon\}.\end{aligned}$$

Il suffit donc de montrer que

$$\overline{G} \ = \ \sum_{i=1}^{k-1} \alpha_i G_i + Q_k Y Q_k^T \ \in \mathcal{C}_n \ .$$

Ainsi, étant donnés les conditions imposées aux α_i, G_i et Y pour que $\overline{G}$ soit dans $\mathcal{G}_{k,\varepsilon}$, notre élément $\overline{G}$ remplira les conditions pour appartenir au sous-différentiel $\partial_\varepsilon \lambda_1(A(x_k))$. Or, pour se trouver dans $\mathcal{C}_n$, il faut être une matrice symétrique s.d.p. et avoir une trace égale à un. $\overline{G}$ remplit ces deux conditions. En effet,

- $\sum_{i=1}^{k-1} \alpha_i G_i + Q_k Y Q_k^T \in \mathcal{S}_n$ et $\succeq 0$.

  Par hypothèse, pour tout $i = 1, \ldots, k-1$, $G_i \in \mathcal{S}_n$ et s.d.p. donc $\alpha_i G_i$ est encore symétrique et s.d.p.. En effet, le vecteur $\widetilde{\alpha}_k$ dont les composantes sont les α_i est dans $\mathbb{R}_+^{k-1}$. D'autre part, suivant la définition de $\mathcal{G}_{k,\varepsilon}$, Y est symétrique et s.d.p. d'ordre r_k. Alors, $Q_k Y Q_k^T$ possède également les propriétés requises. Ainsi, en sommant deux matrices symétriques s.d.p., nous obtenons encore une matrice symétrique s.d.p.
- $\mathrm{Tr}[\sum_{i=1}^{k-1} \alpha_i G_i + Q_k Y Q_k^T] = 1$.

  En utilisant la linéarité de la trace,

$$\begin{aligned} \mathrm{Tr}[\sum_{i=1}^{k-1} \alpha_i G_i + Q_k Y Q_k^T] &= \sum_{i=1}^{k-1} \alpha_i \mathrm{Tr}\ G_i + \mathrm{Tr}\ Q_k Y Q_k^T \\ &= \sum_{i=1}^{k-1} \alpha_i + \mathrm{Tr}\ Y \\ &= 1^T \tilde{\alpha}_k + \langle I_{r_k}, Y \rangle \\ &= 1. \end{aligned}$$

2. Montrons ensuite la première inclusion, i.e. $\delta_{\bar{\varepsilon}} f(x_k) \subset \mathcal{A}^* \mathcal{G}_{k,\varepsilon}$. Par les égalités (2.16) et (2.26), le sous-différentiel approché s'écrit

$$\begin{aligned} \delta_{\bar{\varepsilon}} f(x_k) &= \mathcal{A}^* \delta_{\bar{\varepsilon}} \lambda_1(A(x_k)) \\ &= \mathcal{A}^* \{ Q_{\bar{\varepsilon}} Y Q_{\bar{\varepsilon}}^T : Y \in \mathcal{C}_{r_{\bar{\varepsilon}}} \} \end{aligned}$$

où le $r_{\bar{\varepsilon}}$ représente la dimension de l'espace $E_{\bar{\varepsilon}}(A(x_k))$ et les colonnes de la matrice $Q_{\bar{\varepsilon}}$ forment une base orthonormale de $E_{\bar{\varepsilon}}(A(x_k))$.
En prenant dans $\mathcal{G}_{k,\varepsilon}$, l'élément G^* tel que le vecteur $\widetilde{\alpha}_k$ soit le vecteur nul, i.e. $G^* = Q_k Y Q_k^T$, nous avons bien l'inclusion. En effet, par la définition de $E_{\bar{\varepsilon}}(A(x_k))$, nous avons

$$\langle Q_k Y Q_k^T \ , \ A(x_k) \rangle \geq \lambda_1(A(x_k)) - \bar{\varepsilon} \geq \lambda_1(A(x_k)) - \varepsilon$$

puisque $\varepsilon \geq \bar{\varepsilon}$ par hypothèse. Quel que soit l'élément $\bar{v}$ choisi dans $\delta_{\bar{\varepsilon}} f(x_k)$, il est possible de trouver un élément G^* de $\mathcal{G}_{k,\varepsilon}$ tel que $\bar{v} = \mathcal{A}^* G^*$. $\square$.

L'inclusion (7.1) suggère de prendre l'ensemble $\mathcal{A}^*\mathcal{G}_{k,\varepsilon}$ comme nouvelle approximation de $\partial_\varepsilon f(x_k)$ en x_k.

7.2 Utilisation d'une certaine dualité

Nous allons utiliser la théorie de la dualité pour simplifier notre problème de départ. En faisant référence au problème de projection introduit dans le paragraphe 3.1, ce nouvel ensemble nous mène à un autre problème :

$$\min \|g\|^2 , \quad g \in \mathcal{A}^*\mathcal{G}_{k,\varepsilon} .$$

En posant $\mathcal{G}_k := \mathcal{G}_{k,+\infty}$ et en écrivant un élément de cet ensemble

$$G(\tilde{\alpha}, Y) = \sum_{i=1}^{k-1} \alpha_i G_i + Q_k Y Q_k^T ,$$

nous obtenons un problème de projection équivalent :

$$\begin{cases} \min \|\mathcal{A}^*G(\tilde{\alpha}, Y)\|^2 , \; G(\tilde{\alpha}, Y) \in \mathcal{G}_k \\ f(x_k) - \langle \sum_{i=1}^{k-1} \alpha_i G_i + Q_k Y Q_k^T, A(x_k) \rangle - \varepsilon \leq 0 . \end{cases} \tag{7.2}$$

En utilisant la méthode des pénalités, le problème devient

$$\begin{cases} \min \|\mathcal{A}^*G(\tilde{\alpha}, Y)\|^2 + 2\mu_k (f(x_k) - (\mathcal{A}^*G(\tilde{\alpha}, Y))^T x_k - \langle A_0, G(\tilde{\alpha}, Y) \rangle - \varepsilon) \\ G(\tilde{\alpha}, Y) \in \mathcal{G}_k , \end{cases} \tag{7.3}$$

où μ_k peut être considéré comme un multiplicateur de Lagrange.
Maintenant, en multipliant la fonction objectif par $(\frac{1}{2\mu_k})$ et en laissant tomber les termes constants, nous obtenons le problème suivant

$$\begin{cases} \min \frac{1}{2\mu_k} \|\mathcal{A}^*G(\tilde{\alpha}, Y)\|^2 - (\mathcal{A}^*G(\tilde{\alpha}, Y))^T x_k + \langle A_0, G(\tilde{\alpha}, Y) \rangle \\ G(\tilde{\alpha}, Y) \in \mathcal{G}_k . \end{cases} \tag{7.4}$$

Nous observons que (7.3) est exactement le dual du problème primal

$$\min_{y \in \mathbb{R}^m} \varphi_k(y) + \frac{\mu_k}{2} \|y - x_k\|^2 , \tag{7.5}$$

où $\varphi_k(y)$ est une approximation linéaire de la fonction convexe f.

$$\varphi_k(y) := \max_{G \in \mathcal{G}_k} \langle A(y), G \rangle . \tag{7.6}$$

Cette théorie est présentée de manière plus générale dans [6]. Appliquée à notre problème de départ, nous allons montrer que (7.3) est le dual du problème (7.5). A partir de la proposition 7.1.1, nous pouvons déduire les inclusions suivantes,

$$\{q_k q_k^T\} \subset \mathcal{G}_k \subset \mathcal{C}_n \ , \tag{7.7}$$

où q_k sont des vecteurs propres approchés normalisés associés à $\lambda_1(A(x_k))$. Pour un certain réel ε strictement positif, nous avons alors

$$q_k^T A(x_k) q_k \geq \lambda_1(A) - \varepsilon \ ,$$

et donc, par définition du sous-différentiel approché,

$$q_k q_k^T \in \partial_\varepsilon \lambda_1(A(x_k)) \ .$$

Ainsi, nous sommes assurés que la fonction modèle φ_k possède les propriétés suivantes

$$\begin{cases} \varphi_k(y) \leq f(y) & \forall\, y \in I\!R^n \\ \varphi_k(y_k) \geq f(y_k) - \varepsilon \ . \end{cases} \tag{7.8}$$

Pour résoudre le problème (7.5), un algorithme est présenté dans [6]. Nous ne l'exposerons pas ici. Nous nous limiterons à montrer la dualité entre les deux problèmes. Ecrivons le problème dual de (7.5). Utilisant la définition de notre fonction objectif (7.6) et l'ensemble $\mathcal{G}_k$ étant convexe et borné (car il est inclus dans $\mathcal{C}_n$), nous avons

$$\begin{aligned} \min_{y \in I\!R^n} \varphi_k(y) + \frac{\mu_k}{2}\|y - x_k\|^2 &= \max_{G \in \mathcal{G}_k} \min_{y \in I\!R^n} \langle G(\tilde{\alpha}, Y), A(y) \rangle + \frac{\mu_k}{2}\|y - x_k\|^2 \\ &= \max_{G \in \mathcal{G}_k} F(G) \ . \end{aligned} \tag{7.9}$$

La fonction F atteint son minimum au point qui annule le gradient, par la théorie sur la minimisation sans contrainte différentiable. Ce vecteur qui donne une valeur minimale pour F est

$$y = x_k - \frac{1}{\mu_k}(\mathcal{A}^* G(\tilde{\alpha}, Y)) \ . \tag{7.10}$$

En remplacant y dans (7.9) par sa valeur notée en (7.10), la fonction interne $F(G)$ à maximiser devient

$$\begin{aligned}
F(G) &= \langle G(\tilde{\alpha}, Y), A(x_k - \frac{1}{\mu_k}\mathcal{A}^*G(\tilde{\alpha}, Y))\rangle + \frac{\mu_k}{2}\|\frac{1}{\mu_k}(\mathcal{A}^*G(\tilde{\alpha}, Y))\|^2 \\
&= \langle G(\tilde{\alpha}, Y), A(x_k)\rangle - \frac{1}{\mu_k}\langle G(\tilde{\alpha}, Y), \mathcal{A}(\mathcal{A}^*G(\tilde{\alpha}, Y))\rangle + \frac{1}{2\mu_k}\|\mathcal{A}^*G(\tilde{\alpha}, Y)\|^2 \\
&= \langle G(\tilde{\alpha}, Y), A(x_k)\rangle - \frac{1}{2\mu_k}\|\mathcal{A}^*G(\tilde{\alpha}, Y)\|^2 \\
&= -[\ \frac{1}{2\mu_k}\|\mathcal{A}^*G(\tilde{\alpha}, Y)\|^2 - \langle G(\tilde{\alpha}, Y), A(x_k)\rangle\ ] \\
&= -[\ \frac{1}{2\mu_k}\|\mathcal{A}^*G(\tilde{\alpha}, Y)\|^2 - (\mathcal{A}^*G(\tilde{\alpha}, Y))^T x_k - \langle G, A_0\rangle\ ]\ . \qquad (7.11)
\end{aligned}$$

Nous arrivons donc bien au problème dual décrit en (7.4). □.

Conclusion

De notre analyse découlent deux algorithmes convergents, l'un des *valeurs propres approchées* (premier ordre) et l'autre, du $\mathcal{U}$-*Newton* (second ordre). Pour la mise en place de l'algorithme du premier ordre (algorithme 3.3.1), nous nous sommes appuyées sur la méthode support–boîte noire et sur une recherche dichotomique. La première nous a fournit une direction de descente, direction qui sépare 0 de l'élargissement $\delta_\varepsilon f(x)$ du sous–différentiel $\partial f(x)$. Le long de cette direction, la recherche dichotomique nous a permis de déterminer une longueur de pas. Lorsque la fonction $f := \lambda_1 \circ A$ est bornée inférieurement, nous avons montré que cet algorithme se termine en un nombre fini d'itérations (théorème 3.3.1). Bien que la convergence de ce dernier nécessite peu d'hypothèses, il a été intéressant pour augmenter la précision, d'établir un algorithme du second ordre. Grâce à la condition de transversalité, celui-ci est obtenu en s'appuyant sur l'algorithme du premier ordre ainsi que sur la $\mathcal{U}$-théorie. La convergence de cet algorithme est obtenue de manière semblable à celle de l'algorithme du premier ordre. De plus, en imposant des conditions supplémentaires, la convergence quadratique est atteinte.

Bibliographie

[1] Abraham, R., Marsden, J.E. and Ratiu T., Manifolds, Tensor Analysis, and Applications, *Applied Mathématical Sciences*, 75, second edition, Springer-Verlag, Berlin, 1998.

[2] Bellman, R. and Fan, K., On systems of linear inequalities in Hermitian matrix variables, In Klee, V.L., editor, *Convexité*, volume 7 of *Proceedings of Symposia in Pure Mathematics*, pages 1-11, American Mathematical Society, 1963.

[3] Cullum, J., Donath, W.E. and Balakrishnan, V., The minimization of certain nondifferentiables sums of eigenvalues of symmetric matrices, *Math. Programming Study*, 3 :35-55, 1975.

[4] Edelman, A., Arias, T. and Smith, S.T., Introduction to Minmax, Wiley, 1974.

[5] Fletcher, R., Semi-definite matrix constraints in optimization, *SIAM J. Control Optim.*, 23 :493-523,1995.

[6] Helmberg C. and Oustry, F., Bundle methods to minimize the maximum eigenvalue function, in Vandenberghe, L., Saigal, R. and Wolkovicz, H., editors, *Handbook on Semidefinite Programming. Theory, Algorithms and Applications*, Kluwer Academic Publisher, 2000.

[7] Hicks, N.J., Notes on Differential Geometry, Van Nostrand, Princeton, NJ, 1965.

[8] Hiriart-Urruty, J.B. and Lemaréchal, C., Convex Analysis and Minimization Algorithms, Springer-Verlag, 2 volumes, 1993.

[9] Hiriart-Urruty, J.-B. and Ye, D., Sensitivity analysis of all eigenvalues of a symmetric matrix, *Numerische Mathematik*, 70 :45-72, 1995.

[10] Kiwiel, K. C., A Linearization algorithm for optimizing control systems subject to singular value inequalities, *IEEE Trans Autom. & Control*, AC-31 :595-602, 1986.

[11] Lemaréchal, C., and Oustry, F., Nonsmooth algorithms to solve semidefinite programs, in El Ghaoui, L. and Niculescu, S.-I., editors, *Recent Advances on LMI methods in Control*, Advances in Design and Control series, *SIAM*, 1999 (to appear).

[12] Lemaréchal, C., Oustry, F. and Sagastizabal, C., The $\mathcal{U}$-Lagrangian of a convex function, *Transactions of the American Mathematical Society*, 1997.

[13] Lewis, A. S. and Overton, M. L., Eigenvalue optimization, *Acta Numerica*, 5 :149-190, 1996.

[14] Oustry, F., The $\mathcal{U}$-Lagrangian of the maximum eigenvalue function, *SIAM J. Optimization*, 9(2) :526-549, 1999.

[15] Oustry, F., Vertical developments of a convex function, *Journal of Convex Analysis*, 5(1) :153-170, 1998.

[16] Overton, M.L. and Womersley, R.S., Second derivatives for optimizing eigenvalues of symmetric matrices, *SIAM J. Matrix Anal. Appli.*, 16(3) :667-718, July 1995.

[17] Overton, M.L. and Ye, X., Toward second-order methods for structured nonsmooth optimization, in Gomez, S. and Hennart, J-P., editors, *Advances in Optimization and Numerical Analysis*, pages 97-109, Kluwer Academic Publishers, 1994.

[18] Rockafellar, R.T., Convex Analysis, Princeton University Press, Princeton, NJ, 1970.

[19] Strodiot J.-J., Cours de Programmation Non Linéaire, F.U.N.D.P., 1999.

[20] Strodiot J.-J., Nonsmooth Optimization, Introduction to Numerical Methods, Ho Chi Minh City, December 1999.

[21] Toint P.L., Algèbre, Département de Mathématique F.U.N.D.P., Septembre 1996.

[22] Ye D.Y., Sensitivity analysis of the greatest eigenvalue of a symmetric matrix via the ε-subdifferential of the associated convex quadratic form, *Journal of Optimization Theory and Applications*, 76(2), February 1993.

Annexes

Annexe III.1 *$\mathcal{C}_n$ est un ensemble convexe et compact.*

Preuve.

(i) $\mathcal{C}_n = \{V \in \mathcal{S}_n : V \succeq 0, \mathrm{Tr}(V) = 1\}$ est convexe.
Soient V, V' $\in \mathcal{C}_n$ et $\lambda \in]0,1[$. Montrons que $\lambda V + (1-\lambda)V' \in \mathcal{C}_n$.

1. $\lambda V + (1-\lambda)V' \in \mathcal{S}_n$. En effet, l'ensemble des matrices symétriques est un espace vectoriel.
2. $\lambda V + (1-\lambda)V' \succeq 0$. Puisque nous avons $\lambda > 0$, $(1-\lambda) > 0$ et de plus, V et V' sont semi-défini positifs, nous pouvons conclure.
3. $\mathrm{Tr}(\lambda V + (1-\lambda)V') = 1$ car la trace est linéaire.

(ii) $\mathcal{C}_n$ est compact i.e. fermé et borné.

1. $\mathcal{C}_n$ est fermé.
   Soit $\{V_n\}$ une suite d'élements dans $\mathcal{C}_n$ telle que $V_n \to V$. Vérifions que $V \in \mathcal{C}_n$. En d'autres termes, montrons que pour toute matrice A telle que
   $$\langle V_n, A\rangle \to \langle V, A\rangle,$$
   nous avons $V \in \mathcal{C}_n$. En effet, d'abord $\mathcal{S}_n$ étant fermé et par unicité de la limite, $V \in \mathcal{S}_n$. Ensuite, V est s.d.p. car en choisissant $A = dd^T$ où d est un vecteur non nul, nous avons
   $$d^T V_n d = \langle V_n, dd^T\rangle \to \langle V, dd^T\rangle.$$
   Le membre de gauche étant positif, celui de droite l'est aussi et par conséquent V est s.d.p. Enfin, calculons la trace de V. Prenons pour cela, la matrice A comme la matrice identité, nous avons alors
   $$\langle V_n, I\rangle \to \langle V, I\rangle.$$
   Or, $\mathrm{Tr}(V_n) = 1$, par conséquent $\mathrm{Tr}(V) = 1$.
2. $\mathcal{C}_n$ est borné i.e.
   $$\exists\, M \in I\!R_0^+ \quad \text{tq} \quad \forall\, V \in \mathcal{C}_n\,,\ \langle V, V\rangle \leq M.$$
   Puisque V est une matrice symétrique réelle, nous pouvons considérer sa décomposition spectrale
   $$V = U\Theta U^T$$
   où $\Theta = \mathrm{diag}(\lambda_i(V))$. Dès lors, $V^2 = \Theta^2$ et donc $\mathrm{Tr}(V^2) = \sum_{i=1}^n \theta_i^2$ où θ_i représentent les éléments de la matrice Θ. Et comme $\sum_{i=1}^n \theta_i = 1$, $M = 1$ convient car nous avons
   $$\mathrm{Tr}(V^2) \leq \sum_{i=1}^n \theta_i \;=\; 1\,.$$

□.

Annexe III.2 *L'ensemble $\mathcal{C}_n$ est caractérisé par*

$$\mathcal{C}_n = \text{co}\{qq^T : \ q \in I\!R^n, \ \|q\| = 1\}.$$

Preuve. Nous savons que $\mathcal{C}_n = \{V \in \mathcal{S}_n : V \succeq 0, \ \text{Tr}(V) = 1\}$. Il suffit donc de montrer l'égalité entre les deux ensembles i.e.

$$\{V \in \mathcal{S}_n : V \succeq 0, \ \text{Tr}(V) = 1\} = \text{co}\{qq^T : \ q \in I\!R^n, \ \|q\| = 1\}.$$

(i) Soit $V \in \mathcal{S}_n$ telle que $V \succeq 0$ et $\text{Tr}(V) = 1$. Nous allons montrer que cette matrice V peut s'écrire sous la forme d'une combinaison convexe des qq^T avec $\|q\| = 1$ de sorte que nous obtenions une première inclusion

$$\mathcal{C}_n \subset \text{co}\{qq^T : \ q \in I\!R^n, \ \|q\| = 1\} \ .$$

La matrice V étant une matrice symétrique réelle, nous pouvons considérer sa décomposition spectrale

$$V = Q\Lambda Q^T$$

où Λ est la matrice des valeurs propres de V et Q, la matrice des vecteurs propres orthonormés de V. En considérant le cas $n = 2$, nous avons

$$Q = \begin{bmatrix} q_{11} & q_{21} \\ q_{12} & q_{22} \end{bmatrix}$$

avec q_{ij}, la $j^{ème}$ composante du vecteur propre q_i associé à la $i^{ème}$ valeur propre λ_i de V. Dès lors,

$$\begin{aligned} V &= \begin{bmatrix} q_{11} & q_{21} \\ q_{12} & q_{22} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \\ &= \begin{bmatrix} q_{11} & q_{21} \\ q_{12} & q_{22} \end{bmatrix} \begin{bmatrix} \lambda_1 q_{11} & \lambda_1 q_{12} \\ \lambda_2 q_{21} & \lambda_2 q_{22} \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 q_{11} q_{11} + \lambda_2 q_{21} q_{21} & \lambda_1 q_{11} q_{12} + \lambda_2 q_{21} q_{22} \\ \lambda_1 q_{11} q_{12} + \lambda_2 q_{21} q_{22} & \lambda_1 q_{12} q_{12} + \lambda_2 q_{22} q_{22} \end{bmatrix} \\ &= \lambda_1 \begin{bmatrix} q_{11}^2 & q_{11} q_{12} \\ q_{11} q_{12} & q_{12}^2 \end{bmatrix} + \lambda_2 \begin{bmatrix} q_{21}^2 & q_{21} q_{22} \\ q_{21} q_{22} & q_{22}^2 \end{bmatrix} \\ &= \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T. \end{aligned}$$

Et comme $\sum_{i=1}^{2} \lambda_i = 1$ et $\forall i \ \lambda_i \geq 0$, V est bien de la forme requise.

(ii) Vérifions l'inclusion inverse. Soit $V \in \text{co}\{qq^T : q \in I\!R^n,\ \|q\| = 1\}$. Montrons que $V \in \mathcal{C}_n$. D'abord $V \in \mathcal{S}_n$ car V est une combinaison convexe de matrices symétriques. De plus, $V = \sum_{i=1}^{p} \lambda_i qq^T \succeq 0$ car $qq^T \succeq 0$ pour tout vecteur q, les $\lambda_i \geq 0\ \forall i = 1, \ldots, p$ et une somme de matrices s.d.p. est encore une matrice s.d.p. Enfin, calculons la trace de cette matrice V.

$$\begin{aligned}
\text{Tr}(V) &= \sum_{\text{élts diag.}} \sum_{i=1}^{p} \lambda_i q_i q_i^T \\
&= \sum_{i=1}^{p} \lambda_i \sum_{\text{élts diag.}} q_i q_i^T \\
&= \sum_{i=1}^{p} \lambda_i \sum_{i=1}^{p} \|q_i\|^2 \\
&= \sum_{i=1}^{p} \lambda_i \\
&= 1 \quad \text{par définition de combinaison convexe.}
\end{aligned}$$

Nous obtenons ainsi que $V \in \mathcal{C}_n$. □.

Annexe III.3 *Soient X et $Z \in \mathcal{S}_n$, et $Q = [q_1 \ldots q_r]$ une matrice $n \times r$ dont les colonnes forment une base orthonormale de* $\ker X$.

(i) Alors, $XZ = 0$ si et seulement si
$\exists\ Y \in \mathcal{S}_r$ telle que $Z = QYQ^T$

(ii) Supposons en outre que X et $Z \in \mathcal{S}_n^+$,
alors $\langle X, Z \rangle = 0$ si et seulement si
$\exists\ Y \succeq 0$ telle que $Z = QYQ^T$

Preuve.

(i) Nous avons $XZ = 0 \Leftrightarrow \text{Im}Z \subset \ker X = \text{span}\{q_1, \ldots, q_r\}$. Ce qui est équivalent à dire que Z appartient au sous-espace

$$\text{span}\{q_i q_j^T + q_j q_i^T \quad \forall\ i, j = 1, \ldots, r\} = Q\mathcal{S}_r Q^T. \tag{7.12}$$

Avant de montrer l'égalité (7.12), vérifions l'équivalence suivante

$$\text{Im}Z \subset \ker X \Leftrightarrow Z \in \text{span}\{q_i q_j^T + q_j q_i^T \quad \forall\ i, j = 1, \ldots, r\}.$$

• Examinons la condition nécessaire. Nous prenons le cas particulier où $n = 3$ et $r = 2$; la généralisation est évidente. Nous allons passer par la base canonique, base que nous maîtrisons mieux. Pour y parvenir, nous allons effectuer un changement de base. Soit A une matrice inversible telle que $q_i = Ae_i$ où les e_i sont les vecteurs de la base canonique et les q_i sont des vecteurs qui forment une base de $\ker X$. Nous avons

$$\begin{aligned} Zx &\in \text{span}\{q_1, q_2\} = \text{span}\{Ae_1, Ae_2\} \quad \forall x \in I\!R^3 \\ \text{i.e.} \quad Zx &= \sum_{i=1}^{2} \alpha_i Ae_i \quad \forall x \in I\!R^3 \text{ et } \alpha_i \in I\!R \ . \end{aligned}$$

Par conséquent,

$$A^{-1}Zx \in \text{span}\{e_1, e_2\} \quad \forall x \in I\!R^3 \ ,$$

ou de manière équivalente

$$A^{-1}ZA^{-T}y \in \text{span}\{e_1, e_2\} \quad \forall y \in I\!R^3 \ .$$

En prenant successivement $y = e_1$, $y = e_2$ et $y = e_3$, nous connaissons la forme des colonnes de $A^{-1}ZA^{-T}$, ce qui nous permet d'écrire

$$\begin{aligned} A^{-1}ZA^{-T} &\in \text{span}\{e_1e_2^T + e_2e_1^T, 2e_1e_1^T, 2e_2e_2^T\} \ , \\ \text{i.e.} \quad A^{-1}ZA^{-T} &= \lambda_1(e_1e_2^T + e_2e_1^T) + \lambda_2(2e_1e_1^T) + \lambda_3(2e_2e_2^T) \qquad \text{et } \lambda_i \in I\!R \ . \end{aligned}$$

Donc

$$Z = \lambda_1(Ae_1e_2^TA^T + Ae_2e_1^TA^T) + \lambda_2(2Ae_1e_1^TA^T) + \lambda_3(2Ae_2e_2^TA^T) \ .$$

Par conséquent,

$$Z \in \text{span}\{q_1q_2^T + q_2q_1^T, 2q_1q_1^T, 2q_2q_2^T\}.$$

• Examinons la condition suffisante. Pour tout élément x, nous avons

$$\begin{aligned} Zx &= \sum_{i,j}^{r} b_{ij}(q_iq_j^T + q_jq_i^T)x \quad \text{avec } b_{ij} \in I\!R \ , \\ \text{i.e.} \quad Zx &= \sum_{i,j}^{r} q_ib_{ij}q_j^Tx + \sum_{i,j}^{r} q_jb_{ij}q_i^Tx \ . \end{aligned}$$

En prenant $b_{ij}q_j^Tx = \frac{\alpha_i}{2}$ et $b_{ij}q_i^Tx = \frac{\alpha_j}{2}$ avec $\alpha_i, \alpha_j \in I\!R$, nous obtenons l'inclusion souhaitée.

Maintenant vérifions l'égalité (7.12). Explicitons d'abord le produit matriciel QSQ^T avec $S \in \mathcal{S}_r$ (r=2).

$$\begin{aligned} QSQ^T &= \begin{bmatrix} q_{11} & q_{21} \\ q_{12} & q_{22} \\ q_{13} & q_{23} \end{bmatrix} \begin{bmatrix} s_1 & s_2 \\ s_2 & s_3 \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \end{bmatrix} \\ &= \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \end{aligned}$$

où

$$\begin{aligned} r_{11} &= s_1 q_{11} q_{11} + s_2(q_{21} q_{11} + q_{11} q_{21}) + s_3 q_{21} q_{21} \\ r_{21} &= s_1 q_{11} q_{12} + s_2(q_{22} q_{11} + q_{12} q_{21}) + s_3 q_{22} q_{21} \\ r_{31} &= s_1 q_{11} q_{13} + s_2(q_{23} q_{11} + q_{13} q_{21}) + s_3 q_{23} q_{21} \\ r_{12} &= s_1 q_{11} q_{12} + s_2(q_{21} q_{12} + q_{11} q_{22}) + s_3 q_{21} q_{22} \\ r_{22} &= s_1 q_{12} q_{12} + s_2(q_{22} q_{12} + q_{12} q_{22}) + s_3 q_{22} q_{22} \\ r_{32} &= s_1 q_{13} q_{12} + s_2(q_{23} q_{12} + q_{13} q_{22}) + s_3 q_{23} q_{22} \\ r_{13} &= s_1 q_{11} q_{13} + s_2(q_{21} q_{13} + q_{11} q_{23}) + s_3 q_{21} q_{23} \\ r_{23} &= s_1 q_{12} q_{13} + s_2(q_{22} q_{13} + q_{12} q_{23}) + s_3 q_{22} q_{23} \\ r_{33} &= s_1 q_{13} q_{13} + s_2(q_{23} q_{13} + q_{13} q_{23}) + s_3 q_{23} q_{23} \ . \end{aligned}$$

En séparant les termes dépendants de s_1, s_2 *et* s_3, nous obtenons

$$QSQ^T = \frac{s_1}{2}(q_1 q_1^T + q_1 q_1^T) + s_2(q_2 q_1^T + q_1 q_2^T) + \frac{s_3}{2}(q_2 q_2^T + q_2 q_2^T).$$

A présent, montrons la double inclusion.

1. $\boxed{\text{span}\{q_i q_j^T + q_j q_i^T \quad \forall i, j = 1, \ldots, r\} \subset Q\mathcal{S}_r Q^T.}$

   Soit un élément $a(q_1 q_2^T + q_2 q_1^T)$ appartenant au span avec a un réel arbitraire. En prenant la matrice

   $$S = \begin{bmatrix} 0 & a \\ a & 0 \end{bmatrix},$$

   nous avons

   $$a(q_1 q_2^T + q_2 q_1^T) = QSQ^T.$$

2. $\boxed{Q\mathcal{S}_r Q^T \subset \text{span}\{q_i q_j^T + q_j q_i^T \quad \forall\, i, j = 1, \ldots, r\}.}$

   Prenons un élément de $Q\mathcal{S}_r Q^T$ et montrons que nous pouvons trouver des réels a_{ij} tels que $QSQ^T = \sum_{i,j=1}^{r} a_{ij}(q_i q_j^T + q_j q_i^T)$. En faisant

référence à l'expression du produit QSQ^T avec S une matrice de $\mathcal{S}_r$, il suffit de prendre

$$a_{ij} = \begin{cases} \frac{(S)_{ij}}{2} & \text{quand} \quad i = j \\ (S)_{ij} & \text{sinon.} \end{cases}$$

(ii) Soient X et Z des matrices s.d.p., une conséquence du théorème du produit de Schur, rappelée dans les préliminaires est que $\mathrm{Tr}(XZ) = 0$ est équivalent à $XZ = 0$. Alors, en appliquant le point (i), Z est de la forme QYQ^T et

$$Z = QYQ^T \succeq 0 \Leftrightarrow Y \succeq 0.$$

□.

Annexe III.4 $\delta_\varepsilon\lambda_1(X) = \mathrm{co}\{ee^T : \|e\| = 1\,,\ e \in E_\varepsilon(X)\}$ *est convexe et compact.*

Preuve. L'ensemble $\delta_\varepsilon\lambda_1(X)$, par construction, est convexe. Nous devons donc montrer qu'il est compact, c'est-à-dire fermé et borné. Comme $\{ee^T : \|e\| = 1\,,\ e \in E_\varepsilon(X)\}$ est borné car cet ensemble est contenu dans $\mathcal{C}_n$, il suffit de montrer qu'il est fermé. Nous aurons alors que $\delta_\varepsilon\lambda_1(X)$ est compact puisque l'enveloppe convexe d'un compact est compacte.
Soit $(e_ne_n^T)$ une suite telle que, pour tout n, $\|e_n\| = 1$ et $e_n \in E_\varepsilon(X)$ et supposons que $e_ne_n^T$ converge vers une matrice L. Comme (e_n) est bornée et $E_\varepsilon(X)$ est fermé, il existe une sous-suite telle que $e_{n_k} \longrightarrow e$ avec $\|e\| = 1$ et $e \in E_\varepsilon(X)$. Dès lors, la suite $e_{n_k}e_{n_k}^T$ converge vers ee^T et donc $L = ee^T$ avec $\|e\| = 1$ et $e \in E_\varepsilon(X)$. □.

Annexe III.5

$$\delta_\varepsilon\lambda_1(X) = \{Q_\varepsilon Y Q_\varepsilon^T : Y \in \mathcal{C}_{r_\varepsilon}\} = F_{\mathcal{C}_n}(Q_\varepsilon Q_\varepsilon^T) = \partial\lambda_1(Q_\varepsilon Q_\varepsilon^T) \qquad (7.13)$$

Preuve. La première égalité découle directement du théorème 2.1.1. Vérifions les deux autres égalités. Pour cela, montrons que

$$E_\varepsilon(X) = E_1(Q_\varepsilon Q_\varepsilon^T).$$

Nous savons d'une part que les colonnes de Q_ε, de dimension $n \times r_\varepsilon$, forment une base orthonormale de $E_\varepsilon(X)$. D'autre part, comme $\lambda_1(Q_\varepsilon Q_\varepsilon^T) = 1$ car

$$\max_{\|d\|=1} d^T Q_\varepsilon Q_\varepsilon^T d = \max_{\|d\|=1} \|Q_\varepsilon^T d\|^2 = 1,$$

nous pouvons écrire

$$E_1(Q_\varepsilon Q_\varepsilon^T) = \{q \mid Q_\varepsilon Q_\varepsilon^T q = q\}.$$

Par conséquent, chaque colonne de Q_ε appartient à $E_1(Q_\varepsilon Q_\varepsilon^T)$, et donc les colonnes de Q_ε forment une base orthonormale de $E_1(Q_\varepsilon Q_\varepsilon^T)$. Dès lors,

$$E_\varepsilon(X) = E_1(Q_\varepsilon Q_\varepsilon^T).$$

Nous pouvons alors appliquer le théorème 2.1.1 (i). Nous avons donc les égalités suivantes

$$\begin{aligned} F_{\mathcal{C}_n}(Q_\varepsilon Q_\varepsilon^T) &= \mathrm{co}\{qq^T \mid q \in E_1(Q_\varepsilon Q_\varepsilon^T)\} \\ &= \mathrm{co}\{qq^T \mid q \in E_\varepsilon(X)\} \\ &= \delta_\varepsilon \lambda_1(X). \end{aligned}$$

La dernière égalité est satisfaite grâce au théorème 2.1.1 (ii). □.

Annexe III.6 *Soient $X \in \mathcal{S}_n$ et $U \in \mathbb{R}^{n\times n}$ tel que $U^TU = I_n$. Alors il existe des matrices $n \times n$ $(E_\varepsilon, F_\varepsilon, \Sigma, T)$ telles que*

$$\left\{ \begin{array}{ll} \textit{les colonnes de } E_\varepsilon \textit{ sont des vecteurs unités de } E_\varepsilon(X) & (a) \\ \textit{les colonnes de } F_\varepsilon \textit{ sont des vecteurs unités de } F_\varepsilon(X) & (b) \\ \Sigma \textit{ et } T \textit{ sont diagonales et s.d.p.} & (c) \\ \Sigma^2 + T^2 = I & (d) \\ U = E_\varepsilon \Sigma + F_\varepsilon T. & (e) \end{array} \right. \qquad (7.14)$$

Preuve. Nous allons construire les matrices citées dans le théorème et montrer qu'elles vérifient les cinq conditions. Prenons un vecteur $e \in E_\varepsilon(X)$ et $f \in F_\varepsilon(X)$. Prenons la matrice U définie par

$$U = [u_1, \ldots, u_n] \qquad \text{avec} \quad U^TU = I\ ,$$

et décomposons chaque vecteur u_i sur $E_\varepsilon(X) \oplus F_\varepsilon(X) = \mathbb{R}^n$.
Pour $i = 1, \ldots, n$, nous pouvons écrire

$$u_i = \sigma_i e_i + \tau_i f_i\ ,$$

avec

$$\sigma_i = \|\text{proj}_{E_\varepsilon(X)} u_i\| ,$$

$$e_i = \begin{cases} \frac{\text{proj}_{E_\varepsilon(X)} u_i}{\sigma_i} & \text{si} \quad \sigma_i > 0 \\ e & \text{sinon} , \end{cases}$$

$$\tau_i = \|\text{proj}_{F_\varepsilon(X)} u_i\| ,$$

$$\text{et} \quad f_i = \begin{cases} \frac{\text{proj}_{F_\varepsilon(X)} u_i}{\tau_i} & \text{si} \quad \tau_i > 0 \\ f & \text{sinon} . \end{cases}$$

Construisons alors

$$\begin{aligned} E_\varepsilon &= [e_1, \ldots, e_n] \\ F_\varepsilon &= [f_1, \ldots, f_n] \\ \Sigma &= \text{diag}(\sigma_1, \ldots, \sigma_n) \\ T &= \text{diag}(\tau_1, \ldots, \tau_n). \end{aligned}$$

Les conditions (7.14) sont alors vérifiées. En effet,
(a)-(b) Les colonnes de E_ε et de F_ε sont respectivement des vecteurs unités de $E_\varepsilon(X)$ et de $F_\varepsilon(X)$;
(c) Σ et T sont diagonales (par construction) et s.d.p. (les σ_i et τ_i sont positifs) ;
(d) $\Sigma^2 + T^2 = I$ en effet $\forall i = 1, \ldots, n \quad \sigma_i^2 + \tau_i^2 = \|u_i\|^2 = 1$;
(e) $U = E_\varepsilon \Sigma + F_\varepsilon T$ par la décompositions des vecteurs u_i de la matrice U. □.

Annexe III.7 *Soient* $X \in \mathcal{S}_n$, $(E_\varepsilon, F_\varepsilon, \Sigma, T)$ *satisfaisant (2.18a), (2.18b), (2.18c), (2.18d) et* $\Theta = \text{diag}(\theta_1, \ldots, \theta_n) \in \mathcal{C}_n$. *Alors on a*

$$\begin{cases} E_\varepsilon^T X F_\varepsilon = F_\varepsilon^T X E_\varepsilon = 0 & (a) \\ \lambda_{r_\varepsilon}(X) \leq \langle X, E_\varepsilon \Theta F_\varepsilon^T \rangle \leq \lambda_1(X) & (b) \\ \langle X, F_\varepsilon \Theta F_\varepsilon^T \rangle \leq \lambda_{r_\varepsilon+1}(X) & (c) \\ \text{Tr}(\Sigma \Theta T) \leq [\text{Tr}(T \Theta T)]^{\frac{1}{2}} . & (d) \end{cases} \tag{7.15}$$

Preuve.

a) Montrons que $E_\varepsilon^T X F_\varepsilon = 0$, et de la même manière, nous pouvons obtenir que $F_\varepsilon^T X E_\varepsilon = 0$. Par la définition, nous savons que $E_\varepsilon \oplus F_\varepsilon = \mathbb{R}^n$. Les sous–espaces $E_\varepsilon(X)$ et $F_\varepsilon(X)$ sont donc orthogonaux. De plus, ils sont

$X-$invariants. Vérifions-le pour $E_\varepsilon(X)$ (le raisonnement est identique pour $F_\varepsilon(X)$). Nous devons montrer que

$$X(E_\varepsilon(X)) \subset E_\varepsilon(X) \ .$$

Nous savons par la définition (2.12) qu'un élément de $E_\varepsilon(X)$ s'écrit sous la forme

$$\sum_{i \in I_\varepsilon(X)} a_i e_i \quad \text{avec} \quad a_i \in I\!R^n \text{ et } e_i \in E_i(X) \ .$$

Voyons si

$$X(\sum_{i \in I_\varepsilon(X)} a_i e_i) \in E_\varepsilon(X) \quad \text{avec } a_i \in I\!R^n \text{ et } e_i \in E_i(X)$$

i.e., si

$$\forall i, \ Xe_i \in E_i(X),$$

ou de manière équivalente, si

$$\forall i, \ Xe_i \in \text{span}\{u_j \ : \ Xu_j = \lambda_i u_j\}.$$

Examinons donc un tel élément et fixons i de manière arbitraire. Nous notons $e_i = e$ et $\lambda_i = \lambda$. Nous avons successivement

$$\begin{aligned} Xe &= X \sum_{i \in I_\varepsilon(X)} b_i u_i \\ &= \sum_{i \in I_\varepsilon(X)} b_i X u_i \\ &= \sum_{i \in I_\varepsilon(X)} b_i \lambda u_i \\ &= \lambda \sum_{i \in I_\varepsilon(X)} b_i u_i \\ &\in E_i(X) \end{aligned}$$

par la définition de sous–espace vectoriel pour le i fixé. Donc,

$$Xe_i \in E_i(X) \qquad \forall \, i \in I_\varepsilon(X) \ ,$$

c'est-à-dire $E_\varepsilon(X)$ est X-invariant. Ainsi, les colonnes de XE_ε restent dans $E_\varepsilon(X)$ et sont orthogonales aux colonnes de F_ε.

$$\text{i.e.} \qquad E_\varepsilon^T X F_\varepsilon = 0$$

De manière analogue, nous avons

$$F_\varepsilon^T X E_\varepsilon = 0.$$

b) Puisque

$$E_\varepsilon \Theta E_\varepsilon^T = \sum_{i=1}^{p} \theta_i e_i e_i^T$$

où les e_i sont les colonnes de E_ε, grâce au lemme 2.3.2, nous pouvons écrire

$$\begin{aligned} \langle X, E_\varepsilon \Theta E_\varepsilon^T \rangle &= \langle E_\varepsilon^T X E_\varepsilon, \Theta \rangle \\ &= \sum_{i=1}^{p} \theta_i e_i^T X e_i. \end{aligned}$$

D'autre part, par la définition du sous–espace $E_\varepsilon(X)$ (2.12), tout vecteur unité du sous–espace satisfait

$$\lambda_{r_\varepsilon}(X) \leq e^T X e \leq \lambda_1(x) \ .$$

En effet, vérifions ces deux inégalités .

1. $\boxed{\lambda_{r_\varepsilon}(X) \leq e^T X e}$.

   Tout vecteur e considéré s'écrit

   $$e = \sum_{i \in I_\varepsilon(X)} \alpha_i e_i$$

   avec $\alpha_i \in I\!R^n, e_i \in E_i(X)$ et tel que $\|e\| = 1$. Comme le sous–espace

   $$E_i(X) = \text{span}\{u_j \ : \ X u_j = \lambda_i u_j\} \ ,$$

   nous pouvons écrire

   $$\begin{aligned} Xe &= \sum_{i \in I_\varepsilon(X)} \alpha_i X e_i \\ &= \sum_{i \in I_\varepsilon(X)} \alpha_i \lambda_i e_i. \end{aligned}$$

   Dès lors,

   $$\begin{aligned} e^T X e &= \sum_{i \in I_\varepsilon(X)} \alpha_i e_i^T \alpha_i \lambda_i e_i \\ &= \sum_{i \in I_\varepsilon(X)} \alpha_i^2 \lambda_i \|e_i\|^2 \\ &\geq \lambda_{r_\varepsilon}(X) \sum_{i \in I_\varepsilon(X)} \alpha_i^2 \|e_i\|^2 \ . \end{aligned}$$

Or,

$$\sum_{i \in I_\varepsilon(X)} \alpha_i^2 \|e_i\|^2 = \|e\|^2 = 1 \ .$$

Par conséquent,

$$e^T X e \geq \lambda_{r_\varepsilon}(X) \ .$$

2. $\boxed{e^T X e \leq \lambda_1(X)}$ est vrai car les vecteurs e sont de norme un et

$$\lambda_1(X) = \max_{\|q\|=1} q^T X q.$$

Comme les e_i sont les colonnes de E_ε, alors par (2.18a), nous avons que les vecteurs e_i sont les vecteurs unités de $E_\varepsilon(X)$. Et ainsi,

$$\lambda_{r_\varepsilon}(X) \ \leq \ e_i^T X e_i \ \leq \ \lambda_1(X).$$

Si on multiplie cette dernière ligne par $\sum_{i=1}^n \theta_i = 1$, nous obtenons d'une part

$$\sum_{i=1}^n \theta_i e_i^T X e_i \ \leq \ \sum_{i=1}^n \theta_i \lambda_1(X) \ = \ \lambda_1(X),$$

et d'autre part

$$\sum_{i=1}^n \theta_i e_i^T X e_i \ \geq \ \sum_{i=1}^n \theta_i \lambda_{r_\varepsilon}(X) = \lambda_{r_\varepsilon}(X).$$

En combinant ces deux dernières affirmations, nous obtenons

$$\lambda_{r_\varepsilon}(X) \leq \langle X, E_\varepsilon \Theta E_\varepsilon^T \rangle \leq \ \lambda_1(X).$$

c) Similairement aux inégalités

$$\lambda_{r_\varepsilon}(X) \ \leq \ e^T X e \ \leq \ \lambda_1(X), \quad \forall \, e \in E_\varepsilon(X) \ ,$$

en prenant le complément orthogonal de $E_\varepsilon(X)$, nous obtenons

$$\forall \, f \in F_\varepsilon(X) = (E_\varepsilon(X))^\perp \ : \ \lambda_n(X) \ \leq \ f^T X f \ \leq \ \lambda_{r_\varepsilon+1}(X)$$

puisque $\lambda_{r_\varepsilon+1}$ est la plus grande des valeurs propres restantes. Par conséquent,

$$\begin{aligned} \langle X, F_\varepsilon \Theta F_\varepsilon^T \rangle \ &= \ \sum_{i=1}^p \theta_i f_i^T X f_i \\ &\leq \ \lambda_{r_\varepsilon+1}(X)(\sum_{i=1}^n \theta_i) \\ &= \ \lambda_{r_\varepsilon+1}(X). \end{aligned}$$

d) Par l'égalité (2.18d), nous avons $\Sigma \preceq I_n$. Nous savons également que ΘT est diagonale s.d.p. Par conséquent,

$$\mathrm{Tr}(\Sigma\Theta T) \quad \leq \quad \mathrm{Tr}(\Theta T). \tag{7.16}$$

De plus

$$\begin{aligned} \Theta &= \mathrm{diag}(\theta_1, \ldots, \theta_n) \in \mathcal{C}_n, \\ T &= \mathrm{diag}(t_1, \ldots, t_n) \succeq 0 \end{aligned}$$

et, par la concavité de la racine carrée, nous avons

$$\sum_{i=1}^{n} \theta_i \sqrt{t_i^2} \leq \sqrt{\sum_{i=1}^{n} \theta_i t_i^2}$$

ou encore

$$\mathrm{Tr}(\Theta T) \leq [\mathrm{Tr}(T\Theta T)]^{\frac{1}{2}} .$$

Et en utilisant (7.16), nous obtenons

$$\mathrm{Tr}(\Sigma\Theta T) \leq [\mathrm{Tr}(T\Theta T)]^{\frac{1}{2}}.$$

□.

Annexe III.8 *Soit $X \in \mathcal{S}_n$, $\varepsilon \geq 0$ et $\eta \geq 0$. Alors pour toute matrice $Z \in \partial_\eta \lambda_1(X)$, il existe $G_\varepsilon \in \delta_\varepsilon \lambda_1(X)$ et cinq matrices $n \times n$ $(E_\varepsilon, F_\varepsilon, \Sigma, T, \Theta)$ telles que :*

$$\left\{ \begin{array}{ll} (E_\varepsilon, F_\varepsilon, \Sigma, T) \textit{ satisfont } (2.18a), (2.18b), (2.18c) et (2.18d), & (a) \\ \Theta = \mathrm{diag}(\theta_1, \ldots, \theta_n) \in \mathcal{C}_n & (b) \\ Z = G_\varepsilon + (E_\varepsilon \Sigma\Theta T F_\varepsilon^T + F_\varepsilon \Sigma\Theta T E_\varepsilon^T) + (F_\varepsilon T\Theta T F_\varepsilon^T - E_\varepsilon T\Theta T E_\varepsilon^T) & (c) \\ \mathrm{Tr}(T\Theta T) \leq \frac{\eta}{\lambda_{r_\varepsilon+1}(X) - \lambda_{r_\varepsilon}(X)} = \frac{\eta}{\Delta_\varepsilon(X)} \cdot & (d) \end{array} \right. \tag{7.17}$$

Preuve. Ecrivons la décomposition spectrale de $Z \in \partial_\eta \lambda_1(X) \subset \mathcal{C}_n$: il existe une matrice U de $\mathbb{R}^{n \times n}$ telle que $UU^T = I$ et Θ la matrice diagonale des valeurs propres telles que

$$Z = U\Theta U^T ,$$

ou de manière équivalente,

$$\Theta = U^T Z U .$$

Par conséquent,

$$\mathrm{Tr}(\Theta) = \mathrm{Tr}(Z) = 1.$$

De plus, Z appartenant à $\mathcal{C}_n$, ses valeurs propres sont toutes positives et donc la matrice Θ est s.d.p. Dès lors, $\Theta \in \mathcal{C}_n$, et l'assertion (b) est vérifiée. Appliquons le lemme 2.3.1 qui nous permet d'affirmer l'existence des matrices citées dans (a) qui vérifient

$$U = E_\varepsilon \Sigma + F_\varepsilon T$$

où $(E_\varepsilon, F_\varepsilon, \Sigma, T)$ satisfont (2.18 a-b-c-d). Nous pouvons écrire la décomposition spectrale de Z comme étant

$$Z = (E_\varepsilon \Sigma + F_\varepsilon T)\Theta(E_\varepsilon \Sigma + F_\varepsilon T)^T .$$

En développant le produit matriciel, nous obtenons

$$\begin{aligned} Z &= E_\varepsilon \Sigma \Theta \Sigma^T E_\varepsilon^T + E_\varepsilon \Sigma \Theta T^T F_\varepsilon^T + F_\varepsilon T \Theta \Sigma^T E_\varepsilon^T + F_\varepsilon T \Theta T^T F_\varepsilon^T \\ &= E_\varepsilon \Sigma \Theta \Sigma^T E_\varepsilon^T + E_\varepsilon \Sigma \Theta T F_\varepsilon^T + F_\varepsilon T \Theta \Sigma E_\varepsilon^T + F_\varepsilon T \Theta T F_\varepsilon^T . \end{aligned}$$

Pour satisfaire (c), il suffit que

$$\begin{aligned} E_\varepsilon \Sigma \Theta \Sigma^T E_\varepsilon^T &= E_\varepsilon \Theta E_\varepsilon^T - E_\varepsilon T \Theta T E_\varepsilon^T \quad \text{i.e.} \\ &= E_\varepsilon (\Theta - T\Theta T) E_\varepsilon^T . \end{aligned}$$

Il faut donc vérifier que

$$\Theta - T\Theta T = \Sigma\Theta\Sigma = \Sigma^2\Theta .$$

Puisque $\Sigma^2 + T^2 = I$ i.e. $\Sigma^2 = I - T^2$, nous pouvons écrire

$$\begin{aligned} \Sigma\Theta\Sigma &= (I - T^2)\Theta \\ &= \Theta - \Theta T^2 \\ &= \Theta - T\Theta T. \end{aligned}$$

Dès lors, en posant

$$G_\varepsilon := E_\varepsilon \Theta E_\varepsilon^T = \sum_{i=1}^{p} \theta_i e_i e_i^T$$

où les e_i sont les vecteurs unités de $E_\varepsilon(X)$, nous obtenons

$$Z = G_\varepsilon + (E_\varepsilon \Sigma\Theta T F_\varepsilon^T + F_\varepsilon T\Theta\Sigma E_\varepsilon) + (F_\varepsilon T\Theta T F_\varepsilon^T - E_\varepsilon T\Theta T E_\varepsilon^T)\ . \qquad (7.18)$$

En accord avec (2.19),

$$G_\varepsilon \in \delta_\varepsilon \lambda_1(X).$$

Il reste à montrer la dernière assertion. En prenant le produit scalaire de X avec (7.18), nous obtenons

$$\begin{aligned} \langle X, Z\rangle &= \langle X, G_\varepsilon\rangle \\ &\quad +\langle X, (E_\varepsilon \Sigma\Theta T F_\varepsilon^T + F_\varepsilon T\Theta\Sigma E_\varepsilon)\rangle \\ &\quad +\langle X, (F_\varepsilon T\Theta T F_\varepsilon^T - E_\varepsilon T\Theta T E_\varepsilon^T)\rangle. \end{aligned}$$

Examinons les trois termes de ce produit scalaire :

$$\begin{aligned} \langle X, G_\varepsilon\rangle &= \langle X, E_\varepsilon \Theta E_\varepsilon^T\rangle \\ &\leq \lambda_1(X) \qquad \text{par (2.19b)}\ . \\ \langle X, (E_\varepsilon \Sigma\Theta T F_\varepsilon^T + F_\varepsilon T\Theta\Sigma E_\varepsilon)\rangle &= \langle E_\varepsilon^T X F_\varepsilon + F_\varepsilon^T X E_\varepsilon, \Sigma\Theta T\rangle \\ &= 0 \qquad \text{par (2.19a)}\ . \\ \langle X, (F_\varepsilon T\Theta T F_\varepsilon^T - E_\varepsilon T\Theta T E_\varepsilon^T)\rangle &\leq (\lambda_{r_\varepsilon+1}(X) - \lambda_{r_\varepsilon}(X))\mathrm{Tr}(T\Theta T)\ . \end{aligned}$$

Montrons cette dernière inégalité. Par le lemme 2.3.2,

$$\begin{aligned} \langle X, F_\varepsilon T\Theta T F_\varepsilon^T\rangle &= \langle F_\varepsilon^T X F_\varepsilon, T\Theta T\rangle \\ &\leq \lambda_{r_\varepsilon+1}(X)\mathrm{Tr}(T\Theta T) \\ \text{et}\quad \langle X, E_\varepsilon T\Theta T E_\varepsilon^T\rangle &= \langle E_\varepsilon^T X E_\varepsilon, T\Theta T\rangle \\ &\geq \lambda_{r_\varepsilon}(X)\mathrm{Tr}(T\Theta T). \end{aligned}$$

Les deux inégalités se montrant de manière analogue, nous choisissons de vérifier la première. Ecrivons autrement la partie gauche de cette inégalité, à l'aide du lemme 1.2.1.

$$\begin{aligned} \langle T\Theta T, F_\varepsilon^T X F_\varepsilon\rangle &= \mathrm{Tr}(T\Theta T F_\varepsilon^T X F_\varepsilon) \\ &= \sum_{\text{élts diag.}} (T\Theta T F_\varepsilon^T X F_\varepsilon) \\ &= \sum_{\text{élts diag.}} (T\Theta T)_{ii}(F_\varepsilon^T X F_\varepsilon)_{ii} \\ &= \sum_i t_i^2 \theta_i f_i X f_i^T\ . \end{aligned}$$

Or, dans la preuve de l'annexe III.7, nous avions

$$f_i^T X f_i \leq \lambda_{r_\epsilon+1}(X) .$$

Dès lors,

$$\begin{aligned} \mathrm{Tr}(T\Theta T F_\varepsilon^T X F_\varepsilon) &\leq \lambda_{r_\epsilon+1}(X) \sum_i t_i^2 \theta_i \\ &= \lambda_{r_\epsilon+1}(X) \mathrm{Tr}(T\Theta T). \end{aligned}$$

Le produit scalaire de Z avec X peut donc être majoré comme suit

$$\langle X, Z \rangle \leq \lambda_1(X) + (\lambda_{r_\epsilon+1}(X) - \lambda_{r_\epsilon}(X)) \mathrm{Tr}(T\Theta T).$$

Puisque $Z \in \partial_\eta \lambda_1(X)$, par (2.8), nous avons

$$\langle X, Z \rangle \geq \lambda_1(X) - \eta.$$

En combinant ces deux dernières inégalités, nous obtenons

$$\lambda_1(X) - \eta \leq \lambda_1(X) + (\lambda_{r_\epsilon+1}(X) - \lambda_{r_\epsilon}(X)) \mathrm{Tr}(T\Theta T) .$$

Ainsi,

$$\mathrm{Tr}(T\Theta T) \leq \frac{\eta}{(\lambda_{r_\epsilon}(X) - \lambda_{r_\epsilon+1}(X))} ,$$

ce qui termine la preuve. □.

Annexe III.9 *Soit S une partie convexe compacte de $\mathbb{R}^n$ telle que $0 \notin S$ et soit $g = \mathrm{proj}_S 0$. Si $g \in \mathrm{ri}\ S$, alors*

(a) g est orthogonal à aff S *c-à-d. $\langle g, s - g \rangle = 0$ pour tout $s \in S$.*

(b) $p = \mathrm{proj}_{\{s_1,\ldots,s_k\}} 0 \quad \Leftrightarrow \quad p - g = \mathrm{proj}_{\{s_1,\ldots,s_k\}-g} 0$ où $\{s_1, \ldots, s_k\}$ est une partie finie de S.

Preuve. Nous montrons successivement les deux affirmations.
(a) Soit $s \in S$ et $s \neq g$. Par définition de la projection, nous avons

$$\langle g, s - g \rangle \geq 0 . \tag{7.19}$$

Comme $g \in \mathrm{ri}\ S$, il existe $\delta > 0$ tel que $\bar{B}(g, \delta) \cap \mathrm{aff}\ S \subset S$. Soit $s' = g + \frac{\delta(s-g)}{\|s-g\|}$. Alors

$$\begin{aligned} 2g - s' &\in \bar{B}(g, \delta) \text{ car } \|2g - s' - g\| = \|g - s'\| = \delta , \\ 2g - s' &\in \mathrm{aff}\ S \text{ car } 2g - s' \in \mathrm{aff}\ \{s, g\} . \end{aligned}$$

D'où $2g - s' \in S$ et donc par définition de la projection

$$\langle g, 2g - s' - g \rangle \geq 0$$

$$\text{i.e.} \quad \langle g, g - s' \rangle \geq 0$$

$$\text{i.e.} \quad \langle g, -\frac{\delta(s-g)}{\|s-g\|} \rangle \geq 0$$

$$\text{i.e.} \quad \langle g, s - g \rangle \leq 0 \; . \tag{7.20}$$

De (7.19) et (7.20), il suit que $\langle g, s - g \rangle = 0$.
(b) Soit $\{s_1, \ldots, s_k\}$ une partie finie de S. Alors

$$p = \text{proj}_{\{s_1,\ldots,s_k\}} 0 \;\Leftrightarrow\; \forall\, s \in \text{co}\,\{s_1, \ldots, s_k\} \quad \langle p, s - p \rangle \geq 0$$

$$p - g = \text{proj}_{\{s_1,\ldots,s_k\}-g} 0 \;\Leftrightarrow\; \forall\, s \in \text{co}\,\{s_1, \ldots, s_k\} \quad \langle p - g, s - p \rangle \geq 0 \; .$$

Ces deux lignes sont équivalentes car, en vertu de **(a)**, nous avons

$$\begin{aligned} \langle g, s - p \rangle &= \langle g, s \rangle - \langle g, p \rangle \\ &= \|g\|^2 - \|g\|^2 \\ &= 0 \; . \end{aligned}$$

□.

Annexe III.10 *Pour tout vecteur $x \in \mathbb{R}^m$, le paramètre $\varepsilon(x)$ de (2.30) est plus petit que $\bar{\varepsilon}$. Dès lors,*

$$\delta_{\varepsilon(x)} f(x) \subset \partial_{\bar{\varepsilon}} f(x). \tag{7.21}$$

De plus, si f est bornée inférieurement, on a

$$\begin{aligned} &(i) \qquad \Delta_{\varepsilon(x)}(A(x)) \geq \frac{\bar{\varepsilon}}{n} \\ &\textit{ou} \\ &(ii) \qquad 0 \in \delta_{\varepsilon(x)} f(x) \; . \end{aligned} \tag{7.22}$$

Preuve. Prenons $x \in I\!R^m$ et considérons les deux cas suivants :

1) $\boxed{R_{\bar{\varepsilon}}(x) \neq \emptyset}$

Nous savons d'une part par la remarque 3.3.1 que $\Delta_{\varepsilon(x)}(A(x)) \geq \frac{\bar{\varepsilon}}{n}$. Par conséquent, le point (i) est satisfait. D'autre part, sachant que

$$\lambda_i(A(x)) - \lambda_{i+1}(A(x)) \leq \frac{\bar{\varepsilon}}{n}, \qquad \forall\, i = 1, \ldots, \bar{r}(x) - 1\ ,$$

nous avons

$$\begin{aligned}
\varepsilon(x) &= f(x) - \lambda_{\bar{r}(x)}(A(x)) + \frac{\bar{\varepsilon}}{2n} \\
&= \lambda_1(A(x)) - \lambda_{\bar{r}(x)}(A(x)) + \frac{\bar{\varepsilon}}{2n} \\
&= \lambda_1(A(x)) - \lambda_2(A(x)) + \lambda_2(A(x)) - \lambda_3(A(x)) + \lambda_3(A(x)) \\
&- \ldots - \lambda_{\bar{r}(x)-1}(A(x)) \\
&+ \lambda_{\bar{r}(x)-1}(A(x)) - \lambda_{\bar{r}(x)}(A(x)) + \frac{\bar{\varepsilon}}{2n} \\
&\leq \frac{\bar{\varepsilon}}{n}(\bar{r}(x) - 1) + \frac{\bar{\varepsilon}}{2n} \\
&= (\bar{r}(x) - \frac{1}{2}\ )\ \frac{\bar{\varepsilon}}{n}.
\end{aligned}$$

Or $\bar{r}(x) \leq\ n - 1$ (car dans le cas où $R_{\bar{\varepsilon}}(x) = \emptyset$, $\bar{r}(x) = n$) . Dès lors,

$$\frac{\bar{r}(x) - \frac{1}{2}}{n}\ \leq\ 1 - \frac{3}{2n}\ \leq\ 1\ .$$

Ce qui entraîne que $\varepsilon(x) \leq \bar{\varepsilon}$. Il reste à montrer que

$$\delta_{\varepsilon(x)} f(x) \subset \partial_{\bar{\varepsilon}} f(x).$$

Par (2.29) i.e.,

$$\partial f(x) \subset \delta_{\varepsilon} f(x) \subset \partial_{\varepsilon} f(x),$$

nous avons les inclusions suivantes

$$\delta_{\varepsilon(x)} f(x) \subset \partial_{\varepsilon(x)} f(x) \subset \partial_{\bar{\varepsilon}} f(x),$$

où nous utilisons le fait que $\partial_{\varepsilon} f(x)$ est croissante en ε. En effet, montrons que

$$\exists\ \varepsilon_1,\ \varepsilon_2 \text{ avec } \varepsilon_1 \leq \varepsilon_2 \text{ tel que } \partial_{\varepsilon_1} f(x)\ \subset\ \partial_{\varepsilon_2} f(x).$$

Par définition,

$$\partial_{\varepsilon_1} f(x) = \{s \in I\!R^n \mid f(y) \geq f(x) + \langle s, y - x\rangle - \varepsilon_1\}.$$

Soit $\varepsilon_1 < \varepsilon_2$ et $s \in \partial_{\varepsilon_1} f(x)$, alors

$$\begin{aligned} f(y) &\geq f(x) + \langle s, y - x\rangle - \varepsilon_1 \\ &> f(x) + \langle s, y - x\rangle - \varepsilon_2, \end{aligned}$$

et donc $s \in \partial_{\varepsilon_2} f(x)$. Dès lors,

$$\partial_{\varepsilon_1} f(x) \subset \partial_{\varepsilon_2} f(x).$$

2) $\boxed{R_{\bar{\varepsilon}}(x) = \emptyset}$

Dans ce cas, $\bar{r}(x) = n$ et $\varepsilon(x) = \bar{\varepsilon}$. Comme

$$\partial f(x) \subset \delta_\varepsilon f(x) \subset \partial_\varepsilon f(x),$$

$$\delta_{\varepsilon(x)} f(x) \subset \partial_{\varepsilon(x)} f(x) = \partial_{\bar{\varepsilon}} f(x).$$

D'autre part,

$$\delta_\varepsilon f(x) = \mathcal{A}^* \delta_\varepsilon \lambda_1(A(x))$$

et

$$\delta_\varepsilon \lambda_1(A(x)) = \{Q_\varepsilon Y Q_\varepsilon^T : Y \in \mathcal{C}_{r_\epsilon}\}.$$

De plus, $r_{\varepsilon(x)} = \varepsilon(x)$-multiplicité de λ_1 . Par la remarque 3.3.1, $r_{\varepsilon(x)} = \bar{r}(x)$. Par conséquent

$$\mathcal{C}_{r_{\varepsilon(x)}} = \mathcal{C}_{\bar{r}(x)} = \mathcal{C}_n.$$

Ce qui entraîne que

$$\delta_{\varepsilon(x)} f(x) = \mathcal{A}^*(\mathcal{C}_n).$$

Par le lemme 3.3.1, nous en concluons que $0 \in \delta_{\varepsilon(x)} f(x)$. □.

Annexe III.11 *Considérons les trois définitions citées pour les sous-espaces $\mathcal{U}$ et $\mathcal{V}$. Dans ces conditions,*

(i) $\mathcal{U}_2(\bar{p}) = \mathrm{N}_{\partial f(\bar{p})}(g^0) = \{d \in I\!R^n : \langle g - g^0, d\rangle = 0\ ,\ \forall\ g \in \partial f(\bar{p})\}$,
et est indépendant du choix de $g^0 \in \mathrm{ri}\ \partial f(\bar{p})$

(ii) $\mathcal{U}(\bar{p}) = \mathcal{U}_1(\bar{p}) = \mathcal{U}_2(\bar{p})$.

Preuve. (i) Commençons par montrer que $\mathrm{N}_{\partial f(\bar{p})}(g^0)$, qui est par définition le sous–espace $\mathcal{U}_2(\bar{p})$, a bien la forme décrite. Considérons un g^0 arbitraire dans ri $\partial f(\bar{p})$, vu que $\partial f(\bar{p})$ est un ensemble convexe, nous savons ([8], ch.3, par.5.2) que

$$\mathrm{N}_{\partial f(\bar{p})}(g^0) \;=\; \{d \in I\!R^n \;:\; \langle g - g^0, d\rangle \leq 0 \;,\; \forall\, g \in \partial f(\bar{p})\}\;.$$

Dès lors, nous obtenons que $\mathrm{N}_{\partial f(\bar{p})}(g^0)$ contient $\{d \in I\!R^n \;:\; \langle g-g^0, d\rangle = 0 \;,\; \forall\, g \in \partial f(\bar{p})\}$. Il nous reste donc à montrer l'autre inclusion. A cette fin, considérons $d \in \mathrm{N}_{\partial f(\bar{p})}(g^0)$ et $g \in \partial f(\bar{p})$ arbitraires. Nous savons déjà que $\langle g - g^0, d\rangle \leq 0$, voyons si $\langle g - g^0, d\rangle \geq 0$ est vérifié dans de telles conditions. Supposons que $g - g^0 \neq 0$ (dans le cas contraire nous avons automatiquement $\langle g - g^0, d\rangle = 0$) et posons

$$v \;=\; -\frac{g - g^0}{\|g - g^0\|} \;\in\; \mathcal{V}_1 \cap B(0,1)\;.$$

Puisque $g^0 \in \mathrm{ri}\partial f(\bar{p})$, il existe $\eta > 0$ tel que

$$g^0 + (B(0,\eta) \cap \mathcal{V}_1) \subset \partial f(\bar{p})\;.$$

Or, $\eta v \;\in\; \mathcal{V}_1 \cap B(0,\eta)$ et donc $g^0 + \eta v \;\in\; \partial f(\bar{p})$. Par conséquent,

$$\langle g^0 + \eta v - g^0, d\rangle \leq 0\;,$$

i.e.

$$\begin{aligned}\langle \eta v, d\rangle \;&=\; \langle -\eta\frac{g - g^0}{\|g - g^0\|}, d\rangle \\ &=\; \frac{-\eta}{\|g - g^0\|}\langle g - g^0, d\rangle \;\leq\; 0\;,\end{aligned}$$

avec $\eta > 0$, $\|g - g^0\| > 0$ et donc $\langle g - g^0, d\rangle \geq 0$. Par conséquent,

$$\langle g - g^0, d\rangle = 0.$$

Et donc

$$\mathrm{N}_{\partial f(\bar{p})}(g^0) \subset \{d \in I\!R^n \;:\; \langle g - g^0, d\rangle = 0 \;,\; \forall\, g \in \partial f(\bar{p})\}\;.$$

Il reste à voir que ce résultat est indépendant du choix de g^0 dans ri$\partial f(\bar{p})$. Remplaçons dans l'égalité g^0 par un autre élément γ^0 de ri$\partial f(\bar{p})$. Dans ces conditions,

$$\begin{aligned}\mathrm{N}_{\partial f(\bar{p})}(\gamma^0) \;&=\; \{d \in I\!R^n \;:\; \langle g - \gamma^0, d\rangle = 0 \;,\; \forall\, g \in \partial f(\bar{p})\} \\ &=\; \{d \in I\!R^n \;:\; \langle g, d\rangle = \langle \gamma^0, d\rangle \;,\; \forall\, g \in \partial f(\bar{p})\}\;.\end{aligned}$$

En particulier pour $g = g^0$, nous avons $\langle g^0, d\rangle = \langle \gamma^0, d\rangle$ et donc

$$\begin{aligned} \mathrm{N}_{\partial f(\bar{p})}(\gamma^0) &= \{d \in I\!R^n \ : \ \langle g, d\rangle = \langle g^0, d\rangle \ , \ \forall \ g \in \partial f(\bar{p})\} \\ &= \mathrm{N}_{\partial f(\bar{p})}(g^0) \\ &= \mathcal{U}_2(\bar{p}). \end{aligned}$$

(ii) Maintenant, vérifions l'équivalence des définitions proprement dite. Nous vérifions la double inclusion d'abord pour $\mathcal{U}$ et $\mathcal{U}_2$ et ensuite pour $\mathcal{U}$ et $\mathcal{U}_1$. Nous notons $\mathcal{U}(\bar{p}) = \mathcal{U}$, de même pour $\mathcal{U}_1$ et $\mathcal{U}_2$.

a) Tout d'abord, transformons l'expression de $\mathcal{U}$. Puisque nous savons par les résultats d'analyse convexe que $f'(\bar{p}; d) = \max_{g \in \partial f(\bar{p})} \langle g, d\rangle$, nous en déduisons

$$\begin{aligned} -f'(\bar{p}; -d) &= -\max_{g \in \partial f(\bar{p})} \langle g, -d\rangle \\ &= \min_{g \in \partial f(\bar{p})} \langle g, d\rangle \ . \end{aligned}$$

Ainsi, en remplaçant $f'(\bar{p}; d)$ et $f'(\bar{p}; -d)$ dans la définition de $\mathcal{U}$, nous pouvons écrire

$$\mathcal{U} = \{d \in I\!R^n \ : \ \max_{g \in \partial f(\bar{p})} \langle g, d\rangle = \min_{g \in \partial f(\bar{p})} \langle g, d\rangle\} \ .$$

Grâce à cette nouvelle écriture, nous pouvons en conclure que $\mathcal{U} = \mathcal{U}_2$. En effet, procédons en deux étapes.

1. Montrons la première inclusion.
   Soit $d \in \mathcal{U}$, nous avons

$$\begin{aligned} \max_{g \in \partial f(\bar{p})} \langle g, d\rangle &= \min_{g \in \partial f(\bar{p})} \langle g, d\rangle \\ &\leq \langle g, d\rangle \\ &\leq \max_{g \in \partial f(\bar{p})} \langle g, d\rangle \ , \end{aligned}$$

   et donc

$$\langle g, d\rangle = \langle g^0, d\rangle \qquad \forall \ g \in \partial f(\bar{p}) \ ,$$

   i.e. par (i), $d \in \mathcal{U}_2$.

2. Montrons l'inclusion inverse.
   Considérons $d \in \mathcal{U}_2$, nous avons par (i)

$$\langle g^0, d\rangle = \langle g, d\rangle \qquad \forall \ g \in \partial f(\bar{p}) \ ,$$

et donc

$$\langle g^0, d\rangle \;=\; \max_{g\in\partial f(\bar{p})}\langle g, d\rangle \;=\; \min_{g\in\partial f(\bar{p})}\langle g, d\rangle \;,$$

i.e. $d \in \mathcal{U}$.

b) Afin de prouver que $\mathcal{U}_1 = \mathcal{U}$, nous utiliserons les résultats du point a). En effet, nous montrons d'abord que $\mathcal{U} \subset \mathcal{U}_1$ et ensuite que $\mathcal{U}_1 \subset \mathcal{U}_2$ où nous savons que $\mathcal{U}_2 = \mathcal{U}$.

1. Montrons la première inclusion.
   Considérons $d \in \mathcal{U}$ et $v = \sum_j \lambda_j(g_j - \bar{g}) \;\in\; \mathcal{V}_1$, avec $g_j \in \partial f(\bar{p})$. Nous avons

   $$\langle v, d\rangle \;=\; \sum_j \lambda_j(\langle g_j, d\rangle - \langle \bar{g}, d\rangle) \;=\; 0 \;,$$

   puisque $d \in \mathcal{U}$. Dès lors, $d \in \mathcal{V}_1^{\perp} = \mathcal{U}_1$.
2. Examinons la dernière inclusion.
   Prenons un élément $d \in \mathcal{U}_1 = \mathcal{V}_1^{\perp}$. Par définition, $g - \bar{g} \;\in\; \mathcal{V}_2$ pour tout $g \in \partial f(\bar{p})$ et ainsi,

   $$\langle g - \bar{g}, d\rangle \;=\; 0$$

   ou encore

   $$\langle g, d\rangle \;=\; \langle \bar{g}, d\rangle \qquad \forall\, g \in \partial f(\bar{p}) \;.$$

   En particulier, puisque $g^0 \in \partial f(\bar{p})$,

   $$\langle g^0, d\rangle \;=\; \langle \bar{g}, d\rangle$$

   et donc

   $$\langle g, d\rangle \;=\; \langle g^0, d\rangle \qquad \forall\, g \in \partial f(\bar{p}) \;,$$

   ce qui par (i), signifie que $d \in \mathcal{U}_2$.

Nous avons donc montrer que

$$\mathcal{U} \subset \mathcal{U}_1 \subset \mathcal{U}_2 \;=\; \mathcal{U} \;,$$

et ainsi les trois sous–espaces désignent le même espace que nous notons $\mathcal{U}$.
□.

Annexe III.12 *Soient* $(X,G) \in \mathcal{S}_n \times \mathrm{ri}\partial\lambda_1(X)$. *Alors il existe un* $\eta > 0$ *tel que pour toute matrice* $U \in \mathcal{U}(X) \cap B(0,\eta)$, *l'ensemble des minimisants* $V(X,G;U)$ *soit un singleton :*

$$V(X,G;U) = \{V(U)\} \quad \forall\, U \in B(0,\eta)$$

où $V(.)$ *est l'opérateur défini par*

$$V : \mathcal{U}(X) \cap B(0,\eta) \to \mathcal{V}(X)$$

tel que l'opérateur $\mathcal{U}(X) \cap B(0,\delta) \ni U \longmapsto X + U + V(U)$ *est une paramétrisation* C^∞ *de la surface* $\mathcal{M}_r$.

Preuve. Soit $U \in \mathcal{U}(X), V \in V(X,G;U), W \in \partial\lambda_1(X+U+V) \cap (G + \mathcal{U}(X))$. On sait de (2.6) que

$$\partial\lambda_1(X) = \{Q_1(X) Z Q_1^T(X) : Z \in \mathcal{S}_r^+, \mathrm{Tr} Z = 1\}.$$

Le fait que $W \in \partial\lambda_1(X+U+V)$ entraîne la condition de complémentarité :

$$(\lambda_1(X+U+V) I_n - (X+U+V))W = 0\ . \tag{7.23}$$

En effet,

$$W \in \partial\lambda_1(X+U+V) = \mathrm{co}\{qq^T |\quad q^T q = 1,\ q \in E_1(X+U+V)\}\ .$$

Prenons $W = \sum \alpha_i q_i q_i^T$, avec $\alpha_i \geq 0$ et $\sum_i \alpha_i = 1$. Dès lors,

$$\lambda_1(X+U+V) q_i = (X+U+V) q_i$$

et donc

$$\lambda_1(X+U+V) q_i q_i^T = (X+U+V) q_i q_i^T.$$

Par conséquent, en multipliant par α_i et en sommant, nous obtenons le résultat souhaité . Cette condition de complémentarité entraîne la condition de rang suivante :

$$\mathrm{rg}(\lambda_1(X+U+V) I_n - (X+U+V)) + \mathrm{rg} W \leq n\ . \tag{7.24}$$

En effet par (7.23), $\mathrm{Im}(W) \subset \ker[\lambda_1(X+U+V) I_n - (X+U+V)]$. Donc

$$\mathrm{rg}(W) \leq \dim \ker(\lambda_1(X+U+V) I_n - (X+U+V))\ .$$

Or, par des résultats élémentaires d'algèbre linéaire,

$$\dim(\mathrm{Im}(X)) + \dim(\ker(X)) = n\ ,$$

par conséquent, par la définition de rang,

$$\begin{aligned} \mathrm{rg}(W) &\leq n - \dim(\mathrm{Im}(\lambda_1(X+U+V)I_n - (X+U+V))) \\ &= n - \mathrm{rg}(\lambda_1(X+U+V)I_n - (X+U+V)) \,. \end{aligned}$$

Et donc, nous avons bien la condition de rang (7.24). Remarquons qu'en $U = 0, W = G \in \mathrm{ri}\partial\lambda_1(X)$. Par l'égalité (4.2) de [14], nous savons que

$$\mathrm{ri}\partial\lambda_1(X) = \{Q_1(X)ZQ_1(X)^T : Z \in \mathcal{S}_r, Z \succ 0, \mathrm{Tr}Z = 1\}.$$

Dès lors, nous remarquons que la condition de stricte complémentarité a lieu (cfr. [14], rem 6.6) :

$$\mathrm{rg}(\lambda_1(X)I_n - X) = n - r \quad \text{et} \quad \mathrm{rg}(G) = r \,.$$

Alors par la continuité des valeurs propres, avec les corollaires (4.2.3-(iii)) et (4.2.5), nous avons que $\exists\ \eta > 0$ tel que

$$\mathrm{rg}(\lambda_1(X+U+V)I_n - (X+U+V)) \geq n - r \quad \text{et}\ , \mathrm{rg}(W) \geq r \qquad (7.25)$$

$\forall\ U \in B(0,\eta)$ et $\forall\ (V,G) \in V(X,G;U) \times \partial\lambda_1(X+U+V(X,G;U)) \cap (G + \mathcal{U}(X))$.

Montrons la première inégalité, la seconde se démontrant de manière semblable.

1. En guise de lemme, vérifions d'abord que

   $\exists\ \eta$ tq $\forall\ U \in B(0,\eta)\ , V(U) \qquad \lambda_1(X+U+V(U))$ est de multiplicité $\leq\ r$.

   Pour des questions de facilités, nous notons $V(U) = V$. Supposons par l'absurde que

   $$\forall\ \eta > 0,\ \exists\ U \in B(0,\eta) \text{ et } V(U) \text{ tq } \lambda_1(X+U+V) = \ldots = \lambda_{r+1}(X+U+V) \,.$$

   Prenons $\eta = \frac{1}{n}$ et ainsi, nous construisons une suite de matrices dont la multiplicité de la plus grande valeur propre est au moins $r+1$, i.e.

   $$\lambda_1(X+U_n+V_n) = \ldots = \lambda_{r+1}(X+U_n+V_n) \,.$$

   Quand n tend vers $+\infty$, les égalités sont conservées par la continuité des valeurs propres, et cela donne

   $$\lambda_1(X) = \ldots = \lambda_{r+1}(X) \,,$$

   ce qui contredit le fait que $X \in \mathcal{M}_r$.

2. Montrons maintenant (7.25). Soit $U \in B(0,\eta)$ et $V = V(U)$. Supposons que $\lambda_1(X+U+V)$ soit de multiplicité $s \leq r$. Nous avons alors

$$\dim \ker(\lambda_1(X+U+V)I_n - (X+U+V)) = s \ ,$$

et donc,

$$\begin{aligned} \dim \operatorname{Im}(\lambda_1(X+U+V)I_n - (X+U+V)) &= \operatorname{rg}(\lambda_1(X+U+V)I_n \\ &\quad -(X+U+V)) \\ &= n-s \\ &\geq n-r \ . \end{aligned}$$

En combinant (7.24) avec le fait que $\operatorname{rg}(W) \geq r$, nous avons que

$$\operatorname{rg}(\lambda_1(X+U+V)I_n - (X+U+V)) \leq n-r,$$

et donc par (7.24),

$$n - r + \operatorname{rg}(W) \leq n$$

i.e.,

$$\operatorname{rg}(W) \leq r.$$

Par conséquent, $\exists\ \eta > 0$ tel que

$$\operatorname{rg}(\lambda_1(X+U+V)I_n - (X+U+V)) = n-r \quad \text{et } \operatorname{rg}(W) = r \ ,$$

$\forall\ U \in B(0,\eta)$ et $\forall\ (V,G) \in V(X,G;U) \times \partial\lambda_1(X+U+V(X,G;U)) \cap (G + \mathcal{U}(X))$.

Alors en prenant η suffisamment petit, nous avons pour une matrice $U \in B(0,\eta)$

$$X + U + V(X,G;U) \subset B(X,\delta) \cap \mathcal{M}_r$$

où δ est le rayon introduit au corollaire 4.1.2. Nous appliquons alors ce corollaire dans nos conditions et nous trouvons que

$$\operatorname{proj}_{\mathcal{V}(X)}(U + V(X,G;U)) = V(\operatorname{proj}_{\mathcal{U}(X)}U + V(X,G;U))$$

i.e. $V(X,G;U) = \{V(U)\}$ et l'opérateur $V(.)$ est tel que $X+U+V(U)$ est une paramétrisation C^∞ par la proposition 4.1.2 et le corollaire 4.1.2. □.

Annexe III.13 *Supposons que la transversalité ait lieu en $x \in \mathbb{R}^m$ et prenons un élément $g \in \partial f(x)$. Alors, il existe un G unique dans $\partial\lambda_1(X)$ tq $g = \mathcal{A}^*(G)$.*
De plus, si g est dans $\operatorname{ri}\partial f(x)$, alors G est aussi dans $\operatorname{ri}\partial\lambda_1(X)$.
Et, on a en outre,

$$\begin{aligned} \dim \mathcal{V}^f &= \tfrac{r(r+1)}{2} - 1 \\ \dim \mathcal{U}^f &= m + 1 - \tfrac{r(r+1)}{2}. \end{aligned} \tag{7.26}$$

Preuve. Montrons dans un premier temps que l'opérateur

$$\mathcal{V}(A(x)) \ni V \longmapsto \mathcal{A}^*(V)$$

est non singulier. En effet, prenons $V \in V(A(x))$ telle que $\mathcal{A}^*(V) = 0$ et montrons que $V = 0$. Il suffit pour cela que

$$\forall\ Z \in \mathcal{S}_n \quad \langle Z, V \rangle = 0.$$

Puisque $Z \in \mathcal{S}_n$ et que (T) est satisfaite, nous pouvons écrire cette matrice sous la forme $Z = \mathcal{A}(z) + T$ avec $T \in \mathcal{U}(A(x))$ et $z \in I\!R^n$. Par conséquent,

$$\begin{aligned} \langle Z, V \rangle &= \langle \mathcal{A}(z), V \rangle + \langle T, V \rangle \\ &= \langle z, \mathcal{A}^* V \rangle \\ &= 0\ . \end{aligned}$$

Vérifions maintenant les égalités (7.26). Soit Q_1, une matrice $n \times r$, dont les colonnes forment une base orthonormale de $E_1(A(x))$. L'application $\mathcal{V}(A(x)) \ni V \longmapsto \mathcal{A}^*(V)$ étant non singulière, la dimension de $\mathcal{V}^f(x)$ est identique à la dimension de $\mathcal{V}(A(x))$. Or en utilisant l'égalité (5.8), nous avons également

$$\begin{aligned} \dim \mathcal{V}^f(x) &= \dim\{Q_1 Y Q_1^T\ : Y \in \mathcal{S}_r,\ \mathrm{Tr}(Y) = 0\} \\ &= \dim\{Y\ : Y \in \mathcal{S}_r,\ \mathrm{Tr}(Y) = 0\}\ . \end{aligned}$$

Sachant que la dimension de l'espace des matrices symétriques $r \times r$ vaut $\frac{r(r+1)}{2}$ et que l'ensemble

$$\{Y \in \mathcal{S}_r, \mathrm{Tr}(Y) = 0\}$$

est un hyperplan, nous avons que la dimension de $\mathcal{V}^f(x)$ vaut $\frac{r(r+1)}{2} - 1$. Il est alors évident que la dimension de $\mathcal{U}^f(x)$ est égale à $m + 1 - \frac{r(r+1)}{2}$.
Il reste à montrer que $G \in \partial\lambda_1(X)$ est unique. Supposons que G et $G' \in \partial\lambda_1(A(x))$. En utilisant les égalités (2.4), (2.5) et (5.8), nous avons que $G - G' \in \mathcal{V}(A(x))$. Comme l'application $\mathcal{A}^*$ est injective,

$$\mathcal{A}^*(G - G') = 0 \Leftrightarrow G - G' = 0$$

i.e $G = G'$. □.

Annexe III.14 *Supposons que x^* est une solution de (1) et que (SSOC) a lieu en x^*. Alors*
i) x^* est la solution unique de (1),
ii) pour un $\rho > 0$, il existe un $\alpha > 0$ tel que

$$f(x) \leq f(x^*) + \alpha \quad \Rightarrow \quad x \in B(x^*, \rho),$$

iii) pour un $\rho > 0$, il existe un $\bar{\varepsilon}$ et un δ assez petits tel que l'algorithme 5.8 fournit au moins une itération dans $B(x^*, \rho)$, et toutes les itérations suivantes restent dans $B(x^*, \rho)$.

Preuve.

i) Décomposons un vecteur $d \in \mathbb{R}^m$ arbitraire de la manière suivante : $d = u + v$ avec $u \in \mathcal{U}^f(x^*)$ et $v \in \mathcal{V}^f(x^*)$. Par définition du $\mathcal{U}$-Lagrangien (voir théorème 5.2.1), $f(x^* + d) \geq L^f_{\mathcal{U}}(x^*, 0; u)$. D'où

$$f(x^* + d) \geq L^f_{\mathcal{U}}(x^*, 0; u) = f(x^*) + \frac{1}{2} u^T \nabla^2 L^f_{\mathcal{U}}(x^*, 0; 0) u + o(\|u\|^2).$$

Si d est petit, alors u l'est aussi et par conséquent $f(x^* + d) > f(x^*)$.

ii) Par le point i), $\operatorname{argmin}_{x \in \mathbb{R}^m} f(x)$ est borné ; f a des ensembles niveaux bornés [8] : disons que $f(x) \leq f(x_0) + 1$ implique $\|x - x^*\| \leq M$.
Supposons par l'absurde qu'il existe un $\rho > 0$ et une suite $\{x_k\}_{k \in \mathbb{N}}$ tel que pour $k \geq 1$,

$$f(x_k) \leq f(x^*) + \frac{1}{k} \leq f(x_0) + 1 \text{ et } \|x_k - x^*\| > \rho,$$

alors x_k est borné : $\|x_k - x^*\| \leq M$. Extrayons une sous–suite convergent vers un certain $\hat{x}$ et passons à la limite (en utilisant la continuité de f). On obtient alors

$$f(\hat{x}) \leq f(x^*) \text{ et } \|\hat{x} - x^*\| > \rho$$

i.e. $\hat{x} \neq x^*$. Or par le point i), x^* est le minimum de f, ce qui amène à une contradiction .

iii) Observer que l'algorithme 6.5.1 produit une suite décroissante de valeurs de f : chaque itéré satisfait à $\|x - x^*\| \leq M$. Etant donné $\rho > 0$, prenons $\alpha > 0$ comme en ii) et $\bar{\varepsilon}$ et δ tels que $\bar{\varepsilon} + \delta M \leq \alpha$: de (6.7), au moins le dernier itéré $\bar{x}$ satisfait l'équation $f(\bar{x}) \leq f(x^*) + \alpha$; en effet, (6.7) est valable pour tous les y donc en particulier pour x^*, ce qui nous donne $f(\bar{x}) \leq f(x^*) + \bar{\varepsilon} + \delta \|x^* - \bar{x}\|$ et donc $f(\bar{x}) \leq f(x^*) + \bar{\varepsilon} + \delta M$ et comme $\bar{\varepsilon} + \delta M \leq \alpha$, nous obtenons bien $f(\bar{x}) \leq f(x^*) + \alpha$. Dès lors, par le point ii), $\bar{x} \in B(x^*, \rho)$. Si cela se produit avant l'arrêt, ça se passera à chaque itération suivante. □.

Annexe III.15 *Supposons que* $0 < \bar{\varepsilon} < \Delta_0(A(x^*))$. *Alors, il existe* $\rho_2 > 0$ *tel que pour tout* $x \in B(x^*, \rho_2)$,

$$r_{\bar{\varepsilon}}(x) = \bar{r}(x) = r^*,$$

où $\bar{r}(x)$ *est défini en (3.4) et* $r_{\bar{\varepsilon}}(x) := \dim E_{\bar{\varepsilon}}(X)$.

Preuve. Quand $0 < \bar{\varepsilon} < \Delta_0(A(x^*))$, il est clair que dans (3.4)

$$r_{\bar{\varepsilon}}(x^*) = \bar{r}(x^*) = r^* .$$

En effet, par définition, nous avons

$$\begin{aligned}
\Delta_0(A(x^*)) &= \lambda_r(A(x^*)) - \lambda_{r+1}(A(x^*)) , \\
r^* &= \max\{i \text{ tq } \lambda_i(A(x^*)) = \lambda_1(A(x^*))\} , \\
\bar{r}(x^*) &= \begin{cases} \bar{r} = \min\{r \in R_{\bar{\varepsilon}}(x^*)\} & \text{si } R_{\bar{\varepsilon}}(x^*) \neq \emptyset \\ n & \text{sinon} \end{cases} \\
\text{et} \quad R_{\bar{\varepsilon}}(x^*) &= \{r \ : \ \lambda_r(A(x^*)) - \lambda_{r+1}(A(x^*)) \geq \frac{\bar{\varepsilon}}{n} \} .
\end{aligned}$$

Si $R_{\bar{\varepsilon}}(x^*) = \emptyset$, nous avons directement les égalités annoncées. Nous supposons donc que nous sommes dans le cas où $R_{\bar{\varepsilon}}(x^*) \neq \emptyset$. Tout d'abord, nous avons que r^* est le minimum des éléments de $R_{\bar{\varepsilon}}(x)$. En effet, $r^* \in R_{\bar{\varepsilon}}(x)$ car

$$\lambda_{r^*}(A(x)) - \lambda_{r^*+1}(A(x)) = \Delta_0(A(x^*)) > \bar{\varepsilon} \geq \frac{\bar{\varepsilon}}{n} .$$

De plus, r^* est le minimum. Supposons que $r^* - 1$ soit le minimum, nous aurions alors

$$\lambda_{r^*-1}(A(x^*)) - \lambda_{r^*}(A(x^*)) \geq \frac{\bar{\varepsilon}}{n} > 0 .$$

Or, les r^* premières valeurs propres de $A(x^*)$ sont identiques, ce qui impliquerait que $\bar{\varepsilon}$ soit négatif. De même pour tout autre indice inférieur à r^*. Donc, nous obtenons une première égalité.
D'autre part, $r_{\bar{\varepsilon}}(x^*) = r^*$. En effet,

$$\begin{aligned}
r_{\bar{\varepsilon}}(x^*) &= \dim E_{\bar{\varepsilon}}(A(x^*)) \\
&= \dim \oplus_{i \in I_{\bar{\varepsilon}}(X)} E_i(A(x^*)) \\
I_{\bar{\varepsilon}}(A(x^*)) &= \{i \text{ tq } \lambda_i(A(x^*)) \geq \lambda_1(A(x^*)) - \bar{\varepsilon}\} .
\end{aligned}$$

Les indices de l'ensemble $I_{\bar{\varepsilon}}(X)$ satisfont

$$\begin{aligned} \lambda_i(A(x^*)) &> \lambda_1(A(x^*)) - \Delta_0(A(x^*)) \\ &> \lambda_1(A(x^*)) - \lambda_{r^*}(A(x^*)) + \lambda_{r^*+1}(A(x^*)) \\ &> \lambda_{r^*+1}(A(x^*)) \ . \end{aligned}$$

Cette dernière inégalité n'est vérifiée que pour les indices compris entre un et r^*. Donc l'ensemble des indices $I_{\bar{\varepsilon}}(A(x^*))$ est l'ensemble des naturels strictement positifs inférieurs à r^*. La dimension de l'espace propre approché $E_{\bar{\varepsilon}}(X)$ est égale à celle du sous–espace propre associé à $\lambda_1(A(x^*))$. En d'autres termes,

$$\dim E_{\bar{\varepsilon}}(A(x^*)) \ = \ \dim E_1(A(x^*)) \ = \ r^* \ .$$

Autour de x^*, les résultats sont encore valables, car dans un voisinage de $A(x^*)$, les valeurs propres ont les mêmes propriétés grâce à leur continuité. Les égalités sont conservées dans ce voisinage déterminé par un certain ρ_2 strictement positif. □.